Molecular and phenotypic characterisation of the zebrafish pigmentation gene shady

Lopes, Susana Santos

Award date:
2003

Awarding institution:
University of Bath

Molecular and phenotypic characterisation of the zebrafish pigmentation gene *shady*

Submitted by

Susana Santos Lopes

For the degree of PhD of the University of Bath

2003

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my parents,

Pedro, João and Francisca

Aknowledgements

First, I would like to thank Robert Kelsh, my supervisor, for being so enthusiastic as a scientist and most of all for knowing how to pass that on to his students. Thanks for the many helpful discussions during these years and for becoming a very good friend.

I would like to thank to everybody that collaborated and supported me during this thesis. Special thanks to Andrew Ward and Jonathan Slack for making the Developmental group of the Department of Biology and Biochemistry such a nice place to work.

I am grateful to Laurence Hurst for his work on the phylogenetics of *shady*.

I would like to thank everybody in my Lab 0.76 for their help and friendship. Some members to whom I am particularly grateful have moved on like Angela Pauliny and Stone Elworthy. Thanks also to Kirsten and James Dutton, Tom Carney and Valentina Nikolic. I would also like to mention the Ward Lab including Bill Bennet for their help and for providing a very nice work atmosphere. I regret I could not go out with them more often!

I want to thank all the workers that have been in the fish facility starting with John, then Kim, Lianne and now Richard for taking good care of the fish and making microinjections a more social event.

I am very grateful to Pascal Haffters' team, now Robert Geislers' group for teaching me all about mapping genes when I visited the Max-Plank-Institut fur Entwicklungsbiologie in Tuebingen in 1999.

Thank you to the Fundação para a Ciência e Tecnologia and to the Fundo Social Europeu for funding this project.

I want to thank my friends Virginia Portillo and Federico Dajas-Bailador for all their support during these years. Bath became a very special place after meeting them.

Thank you to my neighbours and friends in Wingfield. Betty, the Sherwoods and now the Kilians. A special thanks goes to Lorraine Cope for taking very good care of Joao while I worked.

A big thank you goes to my Parents and my Grandmothers in Portugal for always believing I could do it. For their support and help and for being close all the time.

To my own family that duplicated during these years with the births of Joao and Francisca. For them goes my deepest gratitude for having shared their mother with this project during all their existence. At last I want to say to Pedro that nothing would have been possible without his support before and after our children were born.

Table of contents

Summary

In mammals, pigment cells (melanocytes) are a biologically and medically important derivative of the neural crest, an embryonic cell population that delaminates from the dorsal neural tube to give rise to many derivatives including: pigment cells, neurons and glia from the PNS, and most head cartilage. Unlike in mammals, fish neural crest cells give rise to three distinct types of pigment cells: melanophores (black), xanthophores (yellow) and iridophores (shiny, silver). These together form the characteristic body pigmentation that underlies camouflage, courtship and communication. We are interested in the mechanisms whereby apparently multipotent neural crest cells generate specific derivative cell-types. I have approached this problem by analysis of one major candidate, *shady (shd)*, for a gene critical for the specification of iridophore fate (Kelsh et al., 1996, Development 123, 369-389). I have used multiple markers to demonstrate that *shd* mutants show defects in only one neural crest derivative (iridophores). Cell-transplantation was used to show that *shd* acts cell-autonomously in the iridophores. I adopted a positional cloning approach to identify the *shd* locus and mapped the gene to LG 17 in collaboration with the Max-Plank-Institut fur Entwicklungsbiologie,Tuebingen, Germany. I then isolated a series of PAC clones from the same genomic region and using PAC microinjection into 2 cell-stage embryos I have shown that *shd* mutant embryos can be significantly rescued by one of these PACs. Sequence analysis of the PACs revealed that only one complete gene was encompassed in the PAC that rescued the *shd* phenotype. *In situ* mRNA expression studies show that this candidate is expressed in a subset of neural crest cells from 30 hpf that show the characteristic distribution of iridoblasts/iridophores at later stages. That this candidate was indeed the *shd* gene was strongly supported by our demonstration that injection of an appropriate morpholino phenocopies the *shd* iridophore phenotype. cDNA sequence followed by phylogenetic analysis revealed that *shd* is the zebrafish orthologue of the mammalian *ALK* (anaplastic lymphoma kinase) gene. The human *ALK* gene encodes a receptor tyrosine kinase that is involved in human oncogenesis but about which little is known of its normal function in development. Finally, we generate testable models of how the *shd/alk* gene might be involved in specification or survival of iridoblasts and suggest interactions with other genes necessary for neural crest development.

Abbreviations

AFLPs	Amplified fragment length polymorphisms
ALCL	Anaplastic Large Cell Lymphoma
ALK	Anaplastic lymphoma kinase
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
bHLH/LZ	basic Helix-loop-Helix/Leucine Zipper
BMP	Bone Morphogenetic Protein
bp	base pair
BSA	bovine serum albumin
CAM	Cell Adhesion Molecule
cDNA	copy DNA
cls	*colourless*
cM	centimorgan
CNS	Central Nervous System
DAB	3,3' Diaminobenzidine
dct	*dopachrome tautomerase*
DDC	Duplication Degeneration Complementation
DIG	Digoxigenin
DNA	dioxyribonucleic acid
dNTPs	deoxy ribonucleoside tri-phosphate
Dom	*Dominant megacolon*
dpf	days post fertilisation
DRG	Dorsal Root Ganglia
DTT	dithiothreitol
ECM	Extra Cellular Matrix

EDN3	Endothelin3
EDNRB	Endothelin Receptor B
ednrb1	endothelin receptor b1
EDTA	ethylene diamine tetra acetate
ENU	Ethylnitrosourea
ERK	Extracellular Regulated Kinase
ESTs	Expressed sequenced tags
FGF	Fibroblast Growth Factor
gch	GTP-cyclohydrolase I
HGMP-RC	Human genome mapping project-resource center
HM	hybridisation mix
hpf	hours post fertilisation
HS	Heat shock
IMT	Inflammatory Myofibroblastic Tumor
IPTG	iso-propyl-β-D-thiogalactopyranoside
IRS	Insulin Receptor Substrate
ITPKA	Inositol 1,4,5 – triphosphate 3-kinase A
IVD	isovaleryl-CoA dehydrogenase
Kb	kilobase
LB	Luria Bertani
LD-PCR	long distance polymerase chain reaction
LG	Linkage group
ls	*lethal spotted*
LTK	Leukocyte tyrosine kinase
MAPK	Mitogenic-activated protein kinase
Mb	megabases

mdk	midkine
Mitf	Microphthalmia
MK	midkine
MQ	milliQ water
mRNA	messenger ribonucleic acid
nac	*nacre*
NC	Neural Crest
NIX	Nucleotide Identify X
NPM	Nucleophosmin
ORF	Open Reading Frame
PAC	P1 artificial chromosome
PBS	phosphate buffered saline
PBT	phosphate buffered saline Tween
PBT	Phosphotyrosine Binding Domain
PCR	polymerase chain reaction
PDGFR	platelet-derived growth factor
pfe	*pfeffer*
PFGE	Pulsed Field Gel Electrophoresis
PI3K	phosphatidylinositol 3'kinase
PNS	Peripheral Nervous System
PTN	pleiotrophin
PTU	phenyl-2-thiourea
RA	Retinoic Acid
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
ros	*rose*

rpm	rotations per minute
RT	reverse transcription
RTK	Receptor Tyrosine Kinase
sal	*salz*
SDS	sodium dodecyl sulphate
sdy	*sandy*
SF	*Steel Factors*
shd	*shady*
$shd^{-/-}$	*shady* homozygous mutants
$shd^{+/-}$	*shady* heterozygous mutants
s^{l}	*piebald lethal*
Sl	*Steel*
SMART	Switching mechanism at 5' end of RNA transcript
spa	*sparse*
SSCPs	Single-strand conformation polymorphisms
SSLPs	Simple sequence length polymorphisms
STS	sequenced-tagged site
TBE	tris borate EDTA
TGFβ	Transforming Growth Factor β
TPM	Tropomyosin
trd	*terminally dull*
TU	Tuebingen
U	units
UTR	Untranslated region
UV	ultraviolet

W	*Dominant spotting*
WT	wild-type
YAC	Yeast artificial chromosome

Chapter 1

General Introduction

1.1 *shady*

shady (*shd*) was the name given to one of the zebrafish pigmentation loci isolated in the Tuebingen mutagenesis screen (Kelsh et al., 1996). Mutants at this locus are characterised by a reduced number of reflecting pigment cells (iridophores), hence the name shady. The *shd* complementation group included 12 mutations that formed a clear phenotypic series (Figure 1.1). In the strongest alleles iridophores were mostly absent. This extreme phenotype lead to the inclusion of *shd* in a subclass of mutations that could be involved in pigment cell specification (Kelsh et al., 1996). The work described in this Thesis, *shd* molecular and phenotypic characterisation (Chapters 4 and 5, respectively) was performed to begin to test this prediction.

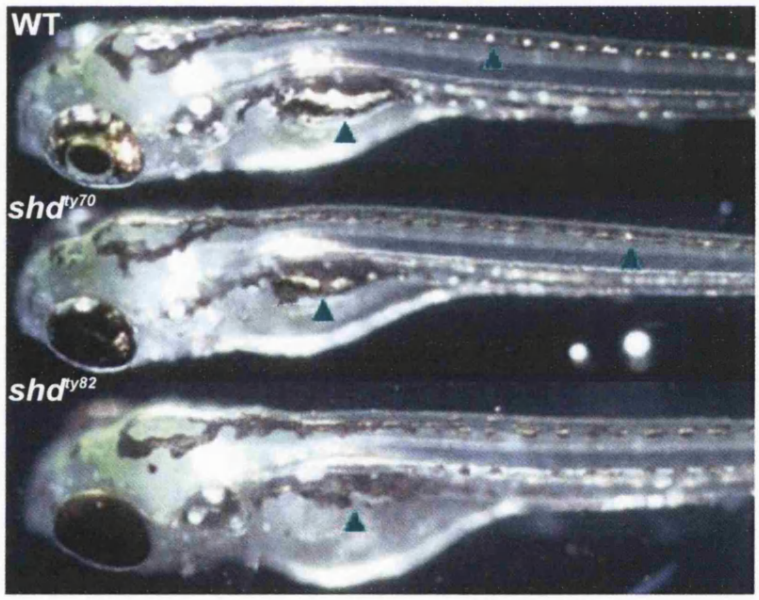


Figure 1.1 *shd* mutants form an allelic series. The strongest allele *shd*^{ty82} has an absence or a severe reduction of iridophores (arrows), while other alleles (e.g. *ty70*) have some iridophores left. The iridophores that remain are normal and look like WT iridophores (Adapted from Kelsh et al. 1996 with permission from The Company of Biologists Ltd.).

All pigment cell types, including iridophores, derive from the neural crest. Therefore, we will introduce the concept of neural crest and explain the complex process of how neural crest cells become pigment cells during vertebrate development. For a better understanding of the relatively recent research done on zebrafish pigment mutants we will introduce some mouse mutants whose molecular characterisation was pivotal in understanding vertebrate pigment cell development.

1.1 Neural Crest development

The vertebrate neural crest is a migratory embryonic cell population that segregates from the tips of the neural folds just before or shortly after they fuse to give rise to the neural tube. Before they emigrate, neural crest cells are an integral part of the neuroepithelium. Neural crest cells undergo an epithelial-mesenchymal transformation, thus generating the neural crest itself. These cells sit transiently as pre-migratory neural crest cells, before becoming motile, enabling migration away from the neural tube. This complex process leads the cells through stereotyped pathways until they reach their final destinations (Eisen and Weston, 1993; LeDouarin and Kalcheim, 1999). How neural crest cells are formed in the first place and how they separate from the neural tube and undergo an epithelial to mesenchymal transformation has been subject of intensive research (Erickson and Perris, 1993; Baker and Bronner-Fraser, 1997; Erickson and Reedy, 1998). In contrast to avian, amphibian and murine embryos the neural tube from teleosts forms from a ventral thickening of the ectoderm that gives rise to a neural keel initially devoid of a central canal (Eisen and Weston, 1993). In teleost fish the neural crest is thought to derive directly from the dorsolateral region of the

neural keel that lies in apposition to the non-neural ectoderm. A similar process in fact occurs in the caudal region of avian embryos (Eisen and Weston, 1993).

To understand how neural crest is generated it was necessary to be able to distinguish these cells from the adjacent ectoderm and the CNS primordium. *Slug*, a zinc finger transcription factor (Nieto et al., 1994) and the cadherin c-Cad-6B (Nakagawa and Takeichi, 1995) were the first premigratory neural crest markers to become available. Using *slug* expression it was possible to recognise cells that formed as the result of an epidermal ectoderm and neural plate contact both *in vivo* and *in vitro*. These cells expressing *slug* were not seen in explants of each separate tissue. As a consequence of this contact between the two tissues other markers were induced in the dorsal region of the neural tube such as *Wnt-1, Wnt-3a, Pax-3 and Dorsalin 1* (Dickinson et al., 1995; Liem et al., 1995). Because some of these responses were dependent on the age of the explanted tissue, it was concluded that the neural tissue has a transient competence to respond to ectodermal signals. This inductive activity of the ectoderm was mimicked *in vitro* by the addition of two distinct members of the TGFβ (transforming growth factor) gene family, BMP4 and BMP7 (Liem et al., 1995). Thus, it was concluded that the generation of the neural crest lineage required active signals from the non-neural ectoderm. It was also investigated whether the neural plate was the only tissue competent to generate neural crest cells. The results indicated that at least until the closure of the neural tube epidermal cells could also form neural crest cells suggesting that there is a bi-directional transmission of inductive signals between the epidermis and the neural plate (Selleck and Bronner-Fraser, 1995). Garcia-Castro et al. (2002) have demonstrated that Wnts and BMPs also play roles in neural crest induction in avians as in amphibians and zebrafish. Wnts induce neural crest from naive neural plates in vitro without added factors whereas BMPs require additives. Therefore these authors believe Wnt molecules are necessary and sufficient to induce neural crest cells in avian embryos (Garcia-Castro et al., 2002).

Aybar et al. (2002) review the induction of neural crest cells and report that as a first step, a gradient of BMPs is established in the ectoderm that results in the segregation into neural plate, neural folds and epidermis at increasing levels of BMP activity. In a second phase, the anterior neural folds are transformed into prospective neural crest by posteriorizing signals from FGF, Wnts and retinoic acid. To finalise the inductive process cell interactions mediated by Notch/Delta signalling are also required (Aybar et al., 2002).

(Aybar and Mayor, 2002) compare the data available on neural crest induction from frog, fish and chick and provide a general common model for this process. The timing of neural crest induction is difficult to determine but some indication can be obtained from the early markers expressed in this tissue, such as *snail, slug, foxD3 and zic.* Expression of these markers indicate that neural crest induction must start soon after the onset of gastrulation in Xenopus and zebrafish. Then, the appropriate level of BMPs (which seem to vary between species) is achieved via anti-BMP molecules, such as noggin. However, because BMPs alone can not induce neural crest (LaBonne and Bronner-Fraser, 1998) it is suggested that Wnts, FGFs, and retinoic acid, which are all expressed at the right time and place in all model systems, have a posteriorizing effect that patterns the anterior-posterior axis of the neural fold.

Once specified, neural crest cells must delaminate from the neural tube to migrate to their destinations. The timing of delamination varies, it happens upon tube closure in avian embryos whereas at this time it is already underway in the head of mouse embryos. In frog and fish neural crest cells wait for several hours before their migration (Nieto, 2001).

Neural crest cells delaminate from the neuroepithelium in a rostro-caudal wave and migrate throughout the embryo using two well defined pathways (dorsoventral and lateral) to give rise to a wide range of derivatives (Figure 1.2) (Baker and Bronner-Fraser, 1997).

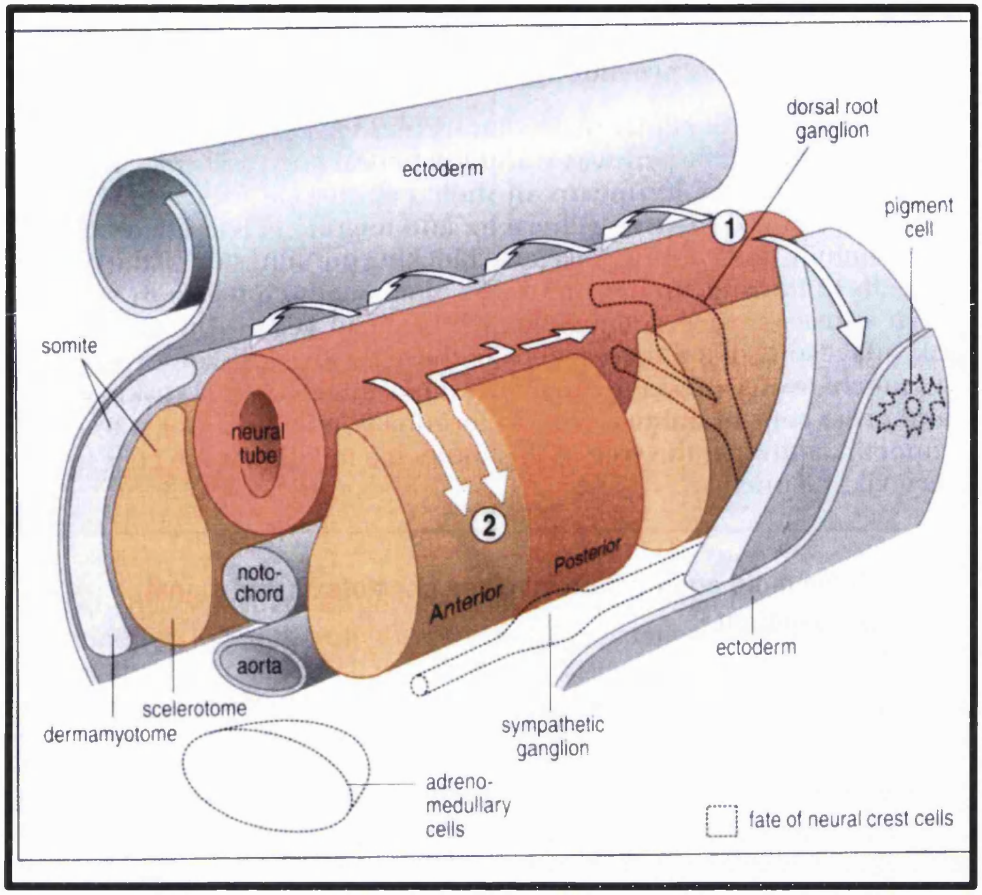


Figure 1.2 Schematic representation of the neural crest migration pathways in the avian embryo. 1 represents the lateral pathway used by melanocytes and 2 indicates the dorsoventral pathway along the anterior part of the somites used by all the other derivatives (Adapted from Wolpert et al, 1998. Principles of Development, Oxford with permission).

How neural crest cells migrate entails dramatic changes in cell-cell and cell-matrix interactions (Erickson and Perris, 1993). Some steps were recognised to be crucial for the neural crest cells to emigrate from the neural tube: First, they appear to be able to breakdown the basal lamina over the dorsal portion of the neural tube (Erickson and Perris, 1993). Second, cell-matrix interactions involving extracellular matrix (ECM) macromolecules such as, fibronectin, laminin, and collagen are known to play an important role for emigrating cells to be able to advance outside the confines of the neural primordium (reviewed by Le Douarin and Kalcheim, 1999). Yet, the molecular basis of this process is largely unknown because the

mode of action of the transmembrane receptors (integrins) that bind these ECM molecules is not yet clearly understood. Third, changes in cell-cell adhesion were found to take place just before the onset of migration of neural crest cells (Newgreen and Gibbins, 1982). These changes involved the loss of intercellular junctions that were both dependent and independent of calcium signalling (Newgreen and Gooday, 1985). Fourth, changes in the plane of cell division were detected in the dorsal neural tube compared to the rest of the neural tube. Around 70% of the mitotic spindles in this region were parallel to the surface of the dorsal tube epithelium meaning that one of the daughter cells could therefore more easily detach and migrate (Erickson and Perris, 1993). To end this list of important steps necessary for migration to occur generation of cell motility has to be added. Several proteins have been involved in neural crest cell motility, members of the TGFβ familly are thought to induce the activation of a cascade of genes (cadherins, cytoskeletal components, transcription factors) that are necessary for crest cell migration (Liem et at., 1995).

In fact, in addition to its role in neural crest induction, BMP4 is also involved in neural crest delamination from the neural tube. While *BMP4* expression is homogeneous along the dorsal neural tube, that of *noggin* is progressively downregulated in a caudorostral manner coinciding in time and axial level with the initial delamination of crest cells (Sela-Donenfeld and Kalcheim, 1999). Various other molecules have been implicated in the epithelial to mesenchymal transition, such as the transcription factor Slug which has been shown to be involved in both the formation and the migration of neural crest (LaBonne and Bronner-Fraser, 2000). Downstream targets of Slug, such as cadherins, have been suggested by Nieto (1994). Recently, Snail, a close relative of Slug, has been shown to repress E-cadherin expression (Cano et al., 2000) making it a good candidate to downregulate N-cadherin or N-CAM expression in the neural crest. Downregulation of some adhesion molecules at the expense of others had been observed at the time of neural crest migration (e.g Nakagawa and Takeichi,

1995). As with Slug, rhoB is also induced by BMP signalling. This molecule is a small GTPase, involved in changes in cell shape through the regulation of actin cytoskeleton and cell adhesion. Inhibition of rhoB prevents neural crest delamination (Liu and Jessell, 1998).

A transcription factor involved in the onset of emigration of neural crest from the neural tube is PAX-3 (Goulding et al., 1991). The mouse mutant *splotch* $^{+/-}$ is characterised by a white patch in the abdomen (Russel, 1947) and represents a deletion in the gene coding for PAX-3 (Kessel and Gruss, 1990; Epstein et al., 1991). These mice have defects in neural tube closure, and severe reduction of many neural crest derivatives such as pigment cells, sympathetic and spinal ganglia, enteric neurons and cardiac structures. Although different studies had shown that *splotch* mutants had a severe defect in emigration of neural crest from the neural tube (Moase and Trasler, 1990; Serbedzija and McMahon, 1997), it was the generation of interspecies chimeras between mice and chick embryos that actually revealed that neural crest cells from *splotch* neural tubes could emigrate normally, when grafted into equivalent positions in chick embryos. In contrast, when the mutant mice neural tubes were grafted into chick lateral plate mesoderm, no emigration occurred compared to control wild-type graftings (Serbedzija and McMahon, 1997). These results suggested that PAX-3 is indeed needed for the interaction between neural crest cells and the somites.

Neural crest was first called ganglion crest, a name that dated from the time of the discovery that the spinal ganglia arise from this material (Horstadius, 1950). Other neural crest derivatives include a variety of cell types: in the head we find that cranial neural crest cells will form the entire facial and hypobranchial skeleton as well as different cell types in the cranial ganglia. In the trunk, there are two major pathways of migration: (Figure 1.2) the dorsoventral pathway between neural tube and somites, that gives rise to sensory and sympathetic ganglia, Schwann cells, and chromaffin cells. The second pathway or lateral pathway, is between the

epidermis and the somites and is only used by melanocytes in avian and mammalian embryos. The study of neural crest cell migration has been best studied in avian embryos, where the quail-chick chimera technique and the HNK-1/NC1 antibody that detects migrating crest cells in whole mount embryos have made this system the first to be well studied (Le Douarin et al., 1993; Le Douarin and Kalcheim, 1999).

A recent important finding concerning how chick neural crest cells may choose between the two well defined pathways was reported by Santiago and Erickson (2002). These authors showed that transmembrane ephrins act as bifunctional guindance cues, first repelling the early migrating neural crest cells from the dorsolateral pathway and then stimulating melanoblasts to migrate into this pathway. They have shown that melanoblasts have several Ephrin B receptors that can be bound by the Ephrin B ligands. However, how neural crest cells can be both attracted and repelled by Ephrins is still not known. Other molecules like F-spondin and semaphorins are suggested to act in a concerted way with Ephrin B ligands to inhibit early migration through the dorsolateral pathway (Santiago and Erickson, 2002).

Zebrafish neural crest is similar to that of other vertebrates but has the advantage of consisting of fewer and larger cells that can be easily observed in living embryos (Figure 1.3). It is possible to mark neural crest cells *in vivo* with a fluorescent tracer and follow their movements in the living specimen with high-resolution fluorescence microscopy (Eisen and Weston, 1993; Raible and Eisen, 1994; Kelsh and Eisen, 2000).

Figure 1.3 Zebrafish neural crest cells in a living embryo. 24 hpf neural crest (nc) cells are dorsal to the neural tube (nt). Lateral view of dorsal tail with anterior to the left and dorsal to the top. (Mf) medial fin; (no) notochord; (mu) muscle cells (with permission from R.N.K.).

Neural crest migration in zebrafish follows the same two pathways that are described for other vertebrates: the medial or dorsoventral pathway, between the somites and notochord, and the lateral pathway, between the somites and the epidermis. Migration on the medial pathway is used by all zebrafish neural crest cell types. Unlike in mammals and birds, many zebrafish pigment cell precursors migrate on the medial pathway. The lateral pathway is used exclusively by pigment cells like in other vertebrates (Eisen and Weston, 1993).

1.3 Pigment cells derive from neural crest

1.3.1 Pigment cell types

Pigment cells include melanophores and melanocytes which are both brown melanin containing cells. The term melanophore is usually employed for 'lower vertebrates' (fish, amphibians and reptiles) whereas melanocyte is used for homeothermic animals or 'higher vertebrates' (Le Douarin and Kalcheim, 1999). Although mammals have only melanocytes, 'lower vertebrates' have up to five other cell-types, but only two are seen in embryonic zebrafish. Xanthophores and erythrophores are coloured bright yellow, orange or red and use carotenoides as the major pigments. They used to be called lipophores because of the fat-soluble nature of carotenoids. Pteridines, concentrated in organelles called pterinosomes, also play an important role in xanthophore pigmentation. Finally, the third type of pigment cells are the iridophores, that form the reflecting units present on the surface of silvery fish (see Figure 1.4). Vertebrate iridophores are present in the irises of some birds and in the skin of amphibians and reptiles but it is in fish that they are more widely spread (Denton and Nicol, 1966) functioning as biological reflectors in their skin (Land, 1972). In fish, iridophores contain guanine platelets each of which is composed by a stack of guanine crystals with intervening cytoplasmic spaces (Figure1.5). Thus, the old name for these cells used to be guanophores (Horstadius, 1950). The platelets are arranged parallel to each other and are all inclined at the same small angle (15^{0}) to the plane of the scales. The spatial arrangement and properties of the guanine crystals are those of ideal light reflectors. When the thickness or the spacing of the crystals within a single stack changes so do their spectral reflective properties (Land, 1972). This effect of changing the structure of the stacks is important because it may explain why during procedures such as paraformaldehyde fixation for in situ hybridisation the iridophores loose their reflective properties.

Figure 1.4 School of marine fish showing their iridescent silvery bodies as a result of the iridophore cells in their skin (Adapted from Doubilet, 1995. light in the sea, New York. With permission).

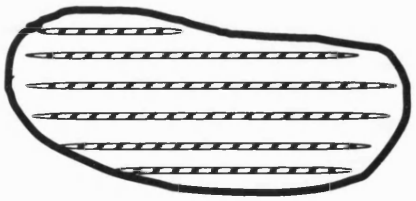


Figure 1.5 Reflectors in fish. Schematic diagram of an electron micrograph section through an iridophore from a silvery fish. The hatched regions are the locations of the broken-up guanine crystals (redrawn from Kawaguti and Kamishima, 1966 *in* Land, 1972).

Iridophores are thought to provide fish with important ecological advantages. The first being the ability to become almost invisible to predators by acting as mirrors that reflect their environment from most points of view. This task is most difficult when the observer is located below the fish because the downwelling light is so much brighter than the upwelling light. For this reason fish developed silvery bellies (covered with iridophores) that act as a "counter-shading" effect (Denton et al., 1972). The second ecological advantage that fish have conferred by iridophores, is being able to signal to their neighbours so as to maintain their positions within a school. By displaying different patterns of bright and dark surfaces depending on the angle of view, such signals can be seen over large distances and provide a basic means of communication (Denton and Rowe, 1994).

1.3.2 Origin of pigment cells

The first suggestions that the neural crest might be the source of pigment cells are due to Borcea back in 1909 (Horstadius, 1950). Harrison (1910, in Horstadius, 1950) found pigment cells in tissue culture from frog spinal cords and suggested that they came from the neural crest but it took another 20 years before more experimental work was done (Horstadius, 1950). Du Shane (1935) was the first to prove that amphibian chromatophores originate from the neural crest. By removing the trunk crest, he found that the ablation of dorsal fin, spinal ganglia, and Rohon-Beard cells were accompanied by the lack of pigment cells (Figure 1.6). Moreover, he found that when grafting pieces of neural folds to the ventral side of an embryo or explanting them *in vitro*, it resulted in the appearance of melanophores.

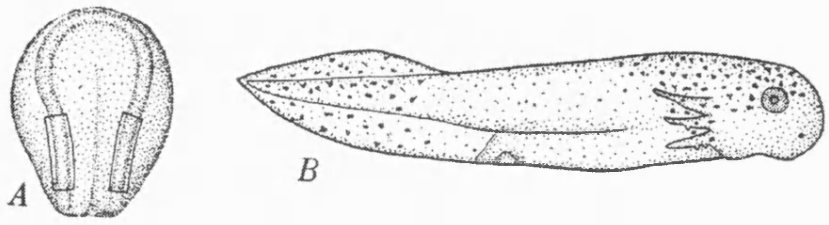


Figure 1.6 Following bilateral excision of trunk neural crest (A) the larva lacks dorsal fin as well as chromatophores in the operated region. Redrawn by Horstadius from original by Du Shane, *J. Exp. Zool. 72,* 1935.

Implantation of crest from frog to urodele showed that also in anurans not only the melanophores, but the iridophores and xanthophores have the same origin (Horstadius, 1950).

In birds, the first experimental evidence for pigment cell origin was found in explants of neural folds in chick, as in cultures they produced branched pigment cells, whereas other tissues or pieces of embryos deprived of neural crest gave no pigment cells (Doris, 1936 in Horstadius, 1950). An interesting experiment that revealed the migratory nature of neural crest cells was one done by Eastlick in 1939 (in Horstadius, 1950). Eastlick transplanted limb-buds in grafts of different size into the coelom or body wall of white chick embryos (Figure 1.7). If the donor was of a pigmented breed and the cut (made after a certain stage) was close to the neural tube, a pigmented leg developed in a white host. But if the graft was cut further away from the neural tube the leg remained unpigmented. In this way the position of the advancing front of the migrating neural crest cells could be established.

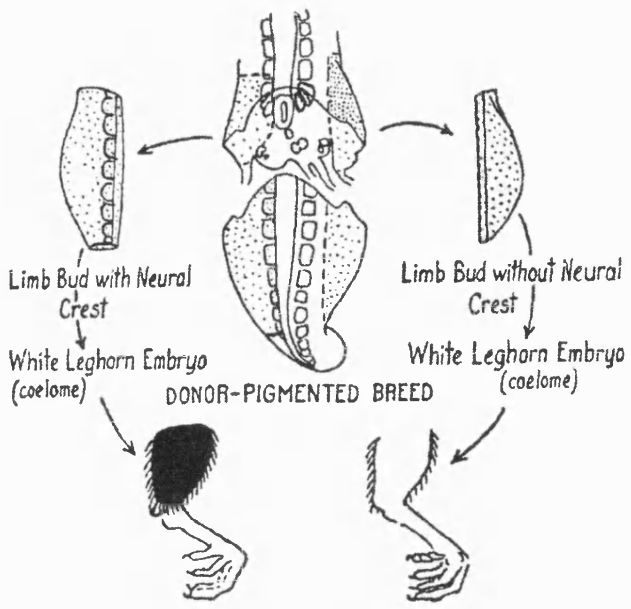


Figure 1.7 Chick limb-buds grafted with (left) or without (right) neural crest cells, depending on the level of the cut in relation to the front of the advancing crest cells. The leg acquires pigment only if the crest cells are included in the graft. Redrawn from Du Shane, *Quart. Rev. Biol. 19*, 1944 by Horstadius, 1950.

In Mammals, only later was there proof of the neural crest origin of melanocytes. Experiments by Rawles (1947, 1948) made use of the fact that mouse embryonic tissues transplanted into the celomic cavity of chick embryos are viable and develop for several days. Explants including the neural crest were taken from black mice and grafted into the celom of unpigmented White Leghorn chicks. It was observed that melanocytes invaded the chick epidermis whereas in control explants, devoid of neural crest, no pigmentation appeared. By transplanting different portions of mouse neural tubes at different stages of development Rawles (1947) also revealed that neural tubes from older mice embryos had lost their pigment-forming capacity at anterior levels.

1.3.3 Chromatoblast theory

The different types of pigment cells referred to above are characterised by the presence of different organelles containing the various pigments. The common origin of pigment cells relies on the observation of mosaic pigment cells such as the melanised iridophores of the dove iris or the erythrophores of some snakes which contain reflective platelets in addition to the usual pterinosomes (Bagnara et al., 1979). These cells contained more than one type of pigment which suggested the presence of a neural crest pigment stem cell containing a primordial organelle of endoplasmic reticular origin that could differentiate in any of the known pigmentary organelles: melanin-containing melanosome; pteridine-containing pterinosome and purine-containing reflecting platelet. Bagnara et al. (1979) believed that such a common neural crest precursor would only be fully committed towards a definite type of chromatophore when it reached its final localisation in the body. Because the chromatophore cell would encounter in its environment different substrates and signals that would provide the appropriate developmental cues to form melanosomes, pterinosomes or reflecting platelets. However, more recent studies from lineage analyses made *in vivo* do not always support this idea. Single neural crest cells from zebrafish originated progeny that consisted of pigment cells alone, neural cells alone, and pigment and neural cells mixed clones. Futhermore, single-derivative clones of each type were also recorded in great numbers (Raible and Eisen, 1994; Dutton et al., 2001). Consequently, a clear proof for the requirement of a common precursor for all pigment cells is not available (see section 1.5 for a more detailed explanation).

1.4 Pigmentation mutants

1.4.1 Some mouse pigmentation mutants

Perhaps the best way to understand the molecular basis of pigmentation is to investigate the pigmentation mutations that are present in nature or that have been generated in animal model systems. Several mouse mutations affecting coat pigmentation have been described and their molecular defects already characterised. We will describe a range of pigment mutant phenotypes that provide important background for understanding the related fish pigmentation mutant phenotypes. For instance, *Dominant spotting* (*W*) and *Steel* (*Sl*) encode for a receptor tyrosine kinase (c-Kit) (Geissler et al., 1988) and its growth factor ligand SF (Steel Factor), (e.g Anderson et al., 1990), respectively. Mice mutant for these genes have a spotted pigmentation pattern, macrocytic anemia and deficiency of primordial germ cells. Mutants for the ligand have a more diluted pigmentation pattern in contrast to the sharp spots of the receptor mutants. The functional significance of the signalling through c-kit in melanoblasts is thought to be proliferation and survival (Yoshida et al., 1996).

The mutants piebald lethal (s^l) and lethal-spotted (ls) are known to encode a G-coupled heptahelical receptor, endothelin receptor B (EDNRB) and its ligand endothelin-3 (EDN3), respectively (Yanagisawa, 1994). Targeted mutations of EDNRB and EDN3 reproduced the phenotypes of the spontaneous mutants (Baynash et al., 1994; Hosoda et al., 1994). Mice mutant for both these genes have decreased melanocytes forming an irregular white spotting pattern, and develop megacolon due to a lack of enteric neurons (also neural crest derived) in the most distal part of the bowel. The absence of the intrinsic reflexes of the enteric nervous system in the terminal gut causes blockage resulting in dilation of the adjacent upstream

colon, i.e. megacolon. Functional studies on the EDN3/EDNRB signalling pathway have shown that it is required for neural crest cell expansion while still in the neural fold, when migrating along the dorso-ventral pathway and as they invade the digestive tract. This signalling is therefore required for the proliferation of the enteric nervous system precursors to colonise the whole gut. In the skin, it is believed that EDN3/EDNRB play a similar role in melanoblasts to the one in the gut (Le Douarin and kalcheim, 1999). Interestingly, Le Douarin and Kalcheim (1999) reviewed important facts that led to the conclusion that EDN3/EDNRB and SF/c-kit might be involved in related but different sequential processes. While the signalling through EDNRB is involved in survival of melanoblasts before they expand, the signalling through c-kit is more likely to be active after the proliferation of these same precursors.

Heterozygous mouse mutants for another gene, *Dominant megacolon* (*Dom*), have phenotypes that are reminiscent of those for *s* and *ls* homozygous mutants, showing reduced number of melanocytes and aganglionosis in the distal part of the gut . *Dom* homozygous mice have also other defects in the PNS outside the enteric nervous system (Pavan and Tilghman, 1994). Molecular characterisation of *Dom* revealed that it encodes the Sox10 transcription factor (Herbarth et al., 1998; Southard-Smith et al., 1998). Mutations in these three genes that generate the terminal aganglionosis of the colon phenotype are models for the human conditions of Hirschsprung's disease and also for the Waardenburg-Shah syndrome, when this phenotype is combined with pigmentation defects (Pingault et al., 1998).

One very important gene in pigmentation is the microphthalmia gene (*Mitf*) that encodes a member of the bHLH-Zip family of transcription factors (Hodgkinson et al., 1993). Most mutant alleles of this gene ablate or severely reduce pigmentation in homozygous mice. *Mitf* mutants often have reduced eye size, early onset deafness, reduced mast and natural killer cell

numbers, and osteoporosis. Homozygous mice have a white coat because they completely lack melanocytes. Thus, this gene seems to be crucial for the development of melanocytes.

Many genes identified in the melanin synthetic pathway were the result of studying mutant mice, such as *silver* and *pink-eyed dilution* which encode proteins involved in the structure of melanosomes; and *Albino, brown* and *slaty* which encode enzymes of the melanogenic pathway (reviewed by Le Douarin and Kalcheim, 1999).

1.4.2 Zebrafish pigmentation mutants

Many pigmentation mutants were isolated during large scale mutagenesis screen both in Tuebingen (Haffter et al., 1996a) and in Boston (Driever et al., 1996) which will be described below, but first it is important to stress the value of such mutations in the discovery of key genes. Mutations causing a homozygous viable phenotype that is easily recognisable are always valuable in genetics and development. Colour is a very visible trait that can be easily scored and therefore pigment mutants such as *Drosophila* with white eyes have had enormous impact on the development of Genetics. Over 90 genes affecting zebrafish pigmentation have been identified in a large-scale genetic-screen in Tuebingen (Haffter et al., 1996b; Kelsh et al., 1996; Odenthal et al., 1996). The ready accessibility of the zebrafish embryo combined with the large size and prominence of their pigment cells allowed comparison of the mutant phenotypes of these genes. This in turn permitted their classification according to the kind of defects they cause, and these suggested which functional role they may have. Kelsh et al. (1996) identified a sub-group of mutations affecting the number of chromatophores present in zebrafish. These mutations are of particular interest because they potentially include genes responsible for the earliest processes of pigment cell development:

specification, proliferation and survival. Defects in these different processes may produce similar phenotypes, as it was noticed in mice before, and it is only through the investigation of their molecular basis that we can dissect the role of these genes.

1.4.2.1 Generation of mutants

The goal of large-scale saturation screens is to identify genes that define developmental pathways. Zebrafish are suitable for such screens because they have a short generation time, high fecundity, rapid development and externally fertilised embryos that are translucent (Haffter and Nusslein-Volhard, 1996). The transparent eggs allow for easy visual and non-invasive detection of embryonic morphological abnormalities. These malformations can easily be classified as being Mendelian or non-Mendelian because of the large number of eggs laid per clutch. During the mutagenesis procedure males are mutagenised with ethylnitrosourea (ENU) and, after the mutations have been fixed, mated to wild-type females. ENU was found to be the most efficient chemical mutagen for inducing point mutations that can be recovered in an F_2 breeding scheme (Mullins et al., 1994). Then the F_1 generation is raised and F_2 families are derived from single-pair matings of F_1 fish. Around 60-80 larvae per F_2 family are raised to adulthood. Half of these fish will be heterozygous for any mutation carried by either of the two F_1 parents. Sibling crosses among F_2 fish will then match two carriers of such a mutation in a quarter of the matings, and a recessive mutant phenotype will be displayed by 25% of the F_3 embryos.

1.4.2.2 Identification of the mutated genes

After generation and detection of the mutant phenotype by visual inspection it is important to investigate further abnormalities that can suggest which developmental pathway has been disrupted. Usually a series of molecular markers are used for careful expression studies.

Often, by phenotypic comparison with mutants present in other model organisms one or more candidate genes might be obvious. On the other hand, phenotypic studies might reveal that the mutated gene is likely to be a completely novel gene revealing a new developmental function. Disruption of such a novel gene can even provide a fish specific phenotype which would result in the absence of obvious candidates from the murine system. This last scenario usually leads to a positional cloning approach as opposed to a candidate approach for isolating the gene in question.

Positional cloning is an unbiased approach that is applicable to any mutation whose inheritance can be traced, even if nothing is known about the gene or biochemical pathways affected by the mutation. Despite having a considerable genome size (two-thirds of mice and human) zebrafish has two advantages for positional cloning projects: the first is the high rate of fecundity that allows the analysis of several thousand meioses and fine mapping of a mutation to a small interval. The second is the accessibility of the embryo which allows for RNA or DNA injection in attempts to rescue the mutant phenotypes and rapid analysis of the expression pattern of candidate cDNAs (Talbot and Schier, 1999). In addition, it is also possible to phenocopy the mutants phenotype by morpholino injection.

Positional cloning projects include three important steps: the first involves the identification of a genetic marker that is near the mutant locus as judged by linkage analysis. The second phase starts when the genetic marker identified in the first step is used to isolate clones from genomic libraries. These clones might contain the gene if the marker is tightly linked to the mutant locus. Otherwise, the process is repeated to carry out a genomic walk until linkage analysis demonstrates that the mutant locus is encompassed in a contiguous stretch of genomic DNA (referred to as the "contig"). The third phase involves the identification of the gene within the contig region. This can be achieved by different procedures, such as

transgenic rescue of the mutation with progressively smaller regions of the contig or sequence analysis of the whole contig (Talbot and Schier, 1999).

1.4.2.3 Some zebrafish pigmentation mutants

According to Kelsh et al. (1996), mutants from the Tuebingen screen that were more likely to be involved in the process of specification included colourless (*cls)* which lacked all three pigment cell types, *salz* and *pfeffer* (*sal, pfe*) with a reduced number of xanthophores (Odenthal et al., 1996), and shady (*shd*) with reduced iridophores (Figure 1.8). Since then, all these mutants, except *shd*, have already been characterised molecularly. The results have shown that specification seems to be a key role for *cls* (Dutton et al., 2001) but not for *pfe,* later named *panther* (Parichy et al., 2000b).

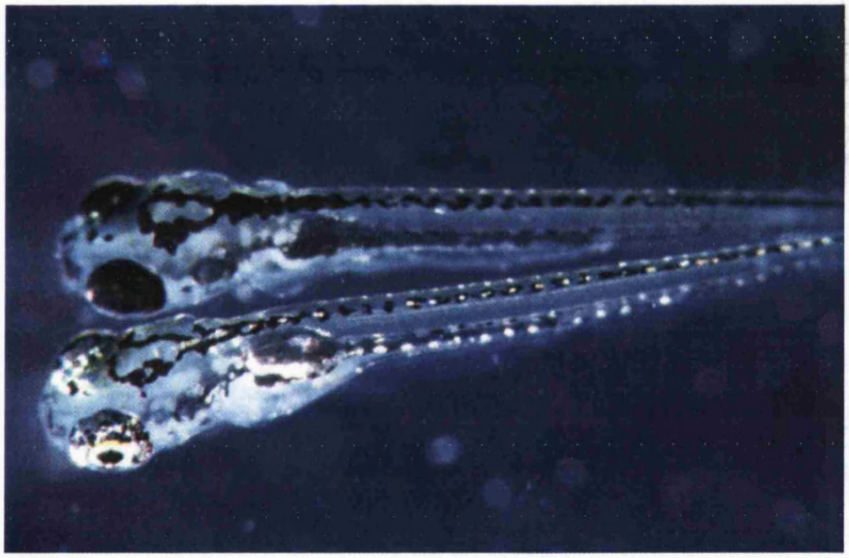


Figure 1.8 *shd* homozygous mutant and wild-type embryo at 6 dpf. *shd* $^{-/-}$ mutants (upper) have very reduced numbers of iridophores all over the eyes, lateral patches and on the dorsal and ventral stripes compared with wild-type (below) siblings.

The pigmentation gene *shd* is an example of a gene that, when mutated, causes a fish specific phenotype: the absence or severe reduction of iridophores (Kelsh et al., 1996). Apart from zebrafish this pigment cell type is only present in one other model organism, *Xenopus laevis*. However, iridophore genes have never been studied in the frog model. Thus, so far, there are no identified genes that could provide a candidate for the *shd* locus. Consequently, *shd* would have to be positional cloned. This formed the basis for much of the work described in this Thesis (see Chapters 2,3,4).

cls, on the other hand, had a phenotype highly reminiscent of the *Dom* homozygous mice where several neural crest derivatives are missing (Herbarth et al., 1998; Southard-Smith et al., 1998). *cls* $^{-/-}$ embryos have fewer melanoblasts as demonstrated by the *dct* riboprobe (Kelsh et al., 2000) and show a drastic reduction in all pigment cell types, enteric neurons and glia, and dorsal root ganglia, (Kelsh and Eisen, 2000). For this reason cloning of the *cls* gene was done by a candidate gene approach which confirmed that *cls* encoded a zebrafish *sox10* homologue to mouse *Sox10* (Dutton et al., 2001). The current model for the role played by *cls* in neural crest development proposes that this gene has a very important role in the specification of all non-ectomesenchymal fates (Dutton et al., 2001).

Zebrafish *sparse* mutants (*spa*) encode the zebrafish *c-kit* (Parichy et al., 1999) orthologue of the mouse *C-kit*. *spa* $^{-/-}$ embryos have a sparse melanophore appearance because of a reduction in melanophore numbers. However, the number of melanoblasts in *spa* mutants is initially normal as revealed by the *dct* marker (Kelsh et al., 2000). Therefore, *spa* function is involved in proliferation and survival of melanophores. Both the phenotype and the proposed function of this gene in fish are thus highly conserved of those in mouse loci *W* and *Sl*. However, *spa/c-kit* function in fish is not essential for hematopoiesis or primordial germ cell development as it is in mouse (Parichy et al., 1999).

A zebrafish mutant phenotype that involves both melanophores and iridophores is the one shown by *nacre* (*nac*) mutants. *nac* encodes the zebrafish basic-helix-loop-helix/leucine zipper transcription factor Mitfa related to the mouse Mitf (Lister et al., 1999). *nac* mutants lack melanophores but interestingly have increased numbers of iridophores (Figure 1.9). Expression of early melanoblast markers like *dct* is absent in *nac* mutants and transplant experiments suggest this gene functions cell autonomously in melanophores. Therefore, the function of *nac* is required for the melanophore fate. Transient expression of the wild-type *nac* gene restored melanophore development in $nac^{-/-}$ mutants suggesting this gene might be sufficient to activate the downstream programme that differentiates melanophores (Lister et al., 1999). More recently, (Elworthy et al., 2003) have shown that this gene can be directly activated by *cls/sox10* and that transient expression of wild-type *nac* is sufficient to rescue melanophore development in both *cls* and *nac* homozygous mutants to the same extent. Taken together the existing data are compelling in demonstrating a key role of *mitfa* in melanophore fate specification.

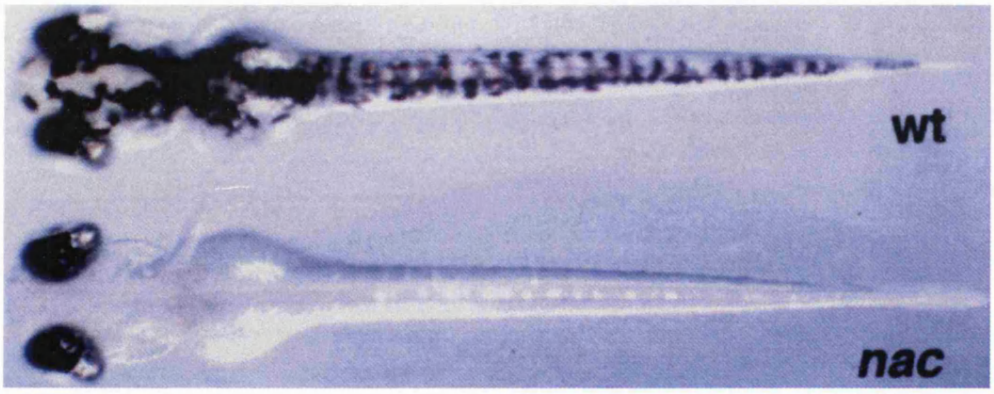


Figure 1.9 *nac* mutants lack melanophores but have increased numbers of iridophores. The chromatophores of the retinal pigmented epithelium that are non crest-derived are normal in $nac^{-/-}$ embryos (with permission from Lister et al.,1999 and The Company of Biologists Ltd.).

Other mutant phenotype that involves both melanophores and iridophores is *rose (ros). ros* mutants encode zebrafish *ednrb1,* an orthologue of the amniote *Ednrb* genes. *ros* mutant

embryos are indistinguishable from wild-types, but adults lack iridophores and also some melanophores. Consequently, the adult stripe pattern is disrupted (Figure 1.10). In spite of the lack of an embryonic phenotype this gene is expressed by precursors for all three embryonic pigment cell types. In the adult fish *ros* is required by a subset of melanophores and by all iridophores. In contrats to murine Ednrb mutants, there is no enteric phenotype. The proposed role for this gene in adult zebrafish is proliferation. The authors suggest that perhaps a paralogous gene might be active in the embryonic pigment cells explaining the lack of an embryonic phenotype (Parichy et al., 2000a).

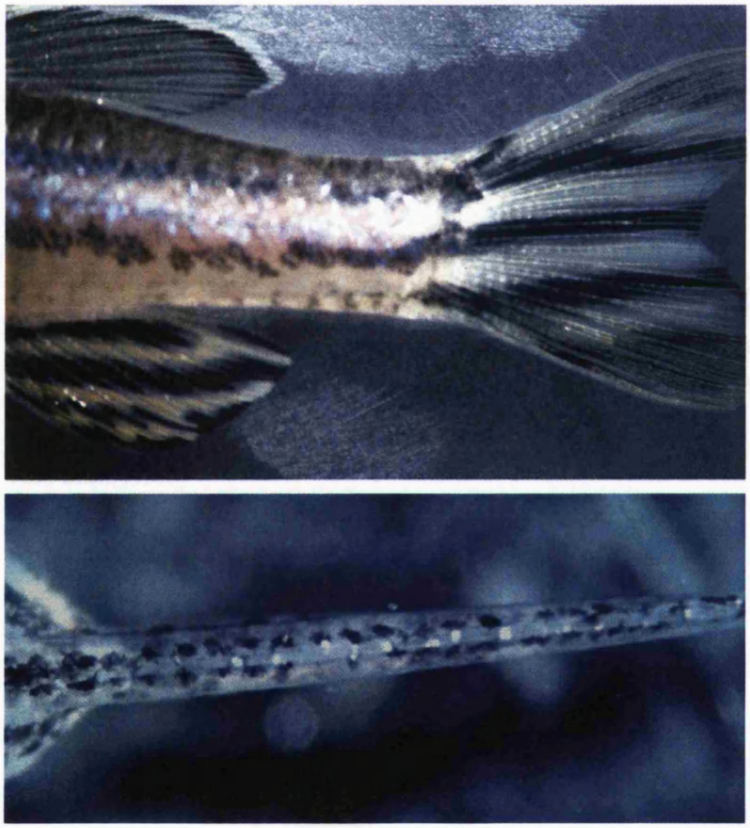


Figure 1.10 *ros* mutant adults lack iridophores and some melanophores (upper) while *ros* mutant embryos (below) have a normal phenotype (from R.N.K. with permission).

It is clear from the majority of the mutant phenotypes described here that there are phenotypic differences between orthologous loci in mouse and fish. In part this lack of similarity results from the fact that fish have more pigment cells types. But surprisingly some differences arise from the absence in fish of phenotypes that are present in the murine mutants. One example is the lack of microphtalmia (small eyes) in fish *nac* $^{-/-}$ embryos compared to the mouse model (Lister et al., 1999). One explanation often given is the sub-functionalization of the zebrafish genome (Force et al., 1999). It is believed that teleosts went through an extra whole genome duplication event. As a consequence, for many genes in mammals there may be two copies in zebrafish (Postlethwait et al., 1999). However, the two genes in zebrafish may be sharing the role of a single gene in mammals, and may have new functions as well. Force et al. (1999) have suggested that duplicate genes may become preserved during evolution by the degenerative loss of complementary gene subfunctions from the duplicate copies. This theory is called the duplication, degeneration, complementation model or DDC. Consistent with this theory, there is another *mitf* orthologue in zebrafish: *mitfb* (Lister et al., 2001). This gene is not expressed in neural crest-derived melanoblasts, in contrast to *mitfa,* but is expressed in the retinal pigmented epithelium (normal in *nac*$^{-/-}$ mutants). m*itfb* is also expressed in other tissues different from the mouse Mitf expression pattern.

1.5 Pigment cell fate specification

Understanding how multipotent neural crest cells give rise to such a diverse range of derivatives is the object of cell fate specification studies. There are two types of studies that try to understand how single neural crest cells can adopt different fates. One is by *in vivo* lineage tracing and the other is by *in vitro* clonal analysis. These procedures approach the question in different ways, *in vivo* studies will provide perhaps more realistic answers because

the fate of individual progenitors is being assessed under the same selective environment as the intact embryo. The *in vitro* experiments, on the other hand, are done under more permissive environments to allow the progenitor cell to reveal its entire range of developmental potentials (Le Douarin and Kalcheim, 1999). Both sets of experiments are very informative and complement each other. A cell becomes specified as soon as it shows a marker for that particular fate. Markers can be gene expression patterns, migration pathways, and final products of differentiation such as pigment.

There are two models for neural crest cell fate restriction: the direct fate restriction model (Bronner-Fraser and Fraser, 1989) where postmigratory pluripotent cells are instructed to adopt a specific fate by environmental signals. The second model is the progressive fate restriction model (Weston, 1982 and Le Douarin et al. 1991) where pluripotent cells undergo a process of gradual restriction.

In vitro clonal analysis of single neural crest cells from quail embryos revealed the presence of three types of clones: 1) clones of pigmented cells only; 2) clones of unpigmented cells that after immunoreactions were identified as sensory and autonomic neurons; 3) mixed clones of pigmented and neuronal progeny. These results suggested the existence of tripotent (melanogenic, sensory and autonomic), bipotent (sensory and autonomic) and monopotent (melanogenic) progenitors that are present at the same time in the trunk neural crest population (reviewed by Sieber-Blum, 1990). Other studies failed to show the existence of the monopotent melanogenic precursor (Baroffio et al., 1988). Instead, it was shown that 20% of clones were indeed composed of a single cell type that was either neuronal, glial, or cartilage. About 80% of the clones were derived from multipotent progenitors with daughter cells composed of 2-4 different types in various combinations. The significance of these results is that they show that although many neural crest cells are pluripotent some progeny apparently

derive from early-specified precursors. This observation was also reported by (Henion and Weston, 1997) when labelling single thoracic neural crest cells growing in culture with a lineage tracer.

The discovery of a truly multipotent cell that could have all neural crest cell types among its progeny (Baroffio et al., 1991) was the reason for the denomination of "neural crest stem cell" in analogy to the stem cells of the hematopoietic system. However, the self-renewal capacity intrinsic to the concept of stem cell was not demonstrated at that time. It was later shown in mammalian embryos that neural crest cells have stem cell properties (Stemple and Anderson, 1992) being able to self-renew up to ten generations. But, these neural crest stem cells appear only to generate a subset of fates, including autonomic neurons, glia, and smooth muscle cells (Kim et al., 2003; Stemple and Anderson, 1992). It is therefore unclear to what extent, and at what time, neural crest stem cells are truly multipotent stem cells, although they can generate 'oligopotent' stem cells, and fate restricted progeny are generated much earlier than previously believed.

The *in vivo* studies using single cell lineage tracing showed that trunk neural crest in birds and in *xenopus* gave rise to multiple types of progenitors thus confirming the existence of a multipotent neural crest precursor. A small proportion of clones was also found to contain only one derivative. In summary, *in vivo* studies re-enforced the *in vitro* findings for these model systems (reviewed by Le Douarin and Kalcheim, 1999).

The same type of experiments conducted in zebrafish showed that neural crest appears to be greatly lineage-restricted in fish compared to other vertebrates. As a model organism for neural crest development zebrafish offers the transparency and accessibility desired for tracking cells in this type of studies. Results from cranial and trunk single cell labelling

experiments show that neural crest cells are more likely to be specified prior to or during migration (Raible and Eisen, 1994; Schilling and Kimmel, 1994). The great majority of the clones analysed by Raible and Eisen (1994) gave rise to only one type of progeny. This was true for all cell types. The combinations of phenotypes in two-derivative clones were never specific combinations of two types. For this reason the authors have no evidence to support the idea that neural crest cells undergo consistent progressive fate restrictions (Raible and Eisen, 1994). An important finding was that in the zebrafish trunk sensory and sympathetic cells are derived from early migrating neural crest cells on the ventral pathway whereas later migrating cells on the lateral pathway would become pigment cells. Also important was the fact that the later the progenitor cells were labelled the higher was the number of clones with single-derivatives (Raible and Eisen, 1994). These data suggest that there is some time/pathway dependent restriction going on. However, it has been demonstrated (Raible and Eisen, 1996) that when early migrating cells are ablated by laser, later migrating neural crest cells also have the plasticity to form sensory and sympathetic neurons and compensate for the lack of an earlier source of these cells. Taken together, these results indicate that zebrafish neural crest cells might be fate-restricted early but are not yet committed. More recently, Dutton et al. (2001) conducted a similar study to Raible and Eisen (1994) and have shown that most of the clones were again of single-derivatives.

Investigation of zebrafish pigment cells in several pigment mutants has also contributed to discussion of cell fate specification. Chromatophores are naturally labelled by the pigment that they contain. Thus, melanin labels melanophores, pteridines label xanthophores and purine platelets and their reflection properties identify iridophores. Some of these natural markers are seen early enough to make chromatophores an advantageous system in which to explore cell fate specification.

At present we know that melanophores are specified prior to migration because *mitfa*, the earliest melanophore marker so far isolated was detected in premigratory neural crest cells (Lister et al. 1999). Other markers such as melanin, *dct* and *c-kit* were seen in migrating cells again indicating that melanophores are already specified as such (Kelsh et al., 2000b; Parichy et al., 1999). In addition, in *cls* mutants pigment cell fate specification seems to largely fail, yet remaining occasional pigment cells occupy a premigratory position suggesting that specification of pigment cells may even be required for normal migration (Dutton et al., 2001; Kelsh and Eisen, 2000).

The work done on *cls* mutants reveals a genetic distinction between ecto and non-ectomesenchymal neural crest fates (Dutton et al., 2001) and suggests that *cls* function may lie high up a cascade of events that specify all pigment cells, neurons and glia from the neural crest. Therefore Kelsh and Raible (2002) propose a model favouring the idea of progressive fate restriction where a non-ectomesenchymal crest cell needs the *cls* gene in addition to a set of specific genes for each fate to become fully specified. In this way, they suggest that crest cells seem to need the action of Wnt signals to specify pigment cell fates at the expense of neural and glial precursors. This was demonstrated by Dorsky et al. (1998) by inhibiting the Wnt signalling pathway and observing neuron formation from cells normally fated to generate pigment cells. Furthermore, melanophores need the action of the *mitfa* transcription factor (Lister et al., 1999) to adopt the melanophore fate. Recent work shows that the activation of the *mitfa* gene can be done directly by *cls/sox10* (Elworthy et al., 2003) and by Wnt signalling (Dorsky et al., 2000). In vivo, the action of both is likely required to specify melanophore fate. Other genes involved in xanthophore and iridophore specification are yet to be discovered.

1.6 Objectives

The principle aim of this project was to identify molecularly the gene encoded by the *shd* locus. The reason to engage in such a project was mainly based on the very specific phenotype of *shd* homozygous embryos suggesting *shd* had a primary role in iridophore specification (Kelsh et al, 1996). We expected that identifying the *shd* gene would provide important tests of whether *shd* is involved in specifying iridophores. In this work we:

1- Localise *shd* in the zebrafish linkage map (Chapter 2)

2- Identify and isolate a short genomic interval containing the *shd* locus (Chapter 3)

3- Identify within this interval candidate genes for the *shd* locus (Chapter 3)

4- Test the best candidate gene by molecular characterisation (Chapter 4)

5- Show that the identified gene is responsible for the *shd* locus (Chapter 4)

6- Investigate the *shd* mutant phenotype (Chapter 5)

7- Provide testable models for *shd* function in the neural crest pigment lineage (Chapter 5 and 6)

Chapter 2

Linkage mapping of the *shady* locus

2.1 Introduction

Genes identified in mutant screens can be cloned in two distinct ways. In some cases, candidate genes can be suggested based on studies in other model organisms. This approach has been used in many zebrafish studies (e.g. Lister et al., 1999; Moens et al., 1998; Parichy et al., 2000a; Parichy et al., 1999) and it has also been used for *cls* in our lab (Dutton et al., 2001). However, where no candidate genes exist, or where the candidates fail to show tight linkage to the mutant phenotypes, positional cloning techniques must be used. This approach was used successfully to elucidate the molecular basis of Duchenne muscular dystrophy (Hoffman, 1987) and has been used to clone many human disease genes (Berger et al., 1992; Boldog et al., 1993; Clark and Goate, 1993; Collins, 1990; Collins, 1991; Douar et al., 1994; Goldberg and Collins, 1991; Ranta et al., 2000). This alternative method has been increasingly used over the last five years to identify mutant loci in zebrafish (Brownlie et al., 1998; Zhang et al., 1998); Talbot and Schier, 1999; Inoue et al., 1999; Andrews, 2000; Kupperman et al., 2000; Guo et al., 2000; Donovan et al., 2000; Rottbauer et al., 2001; Topczewski et al., 2001; Horne-Badovinac et al., 2001; Clark et al., 2001; Jessen et al., 2002; Lyons et al., 2002; Poss et al., 2002; Parsons et al., 2002; Shafizadeh et al., 2002; Xu et al., 2002). This same approach will be necessary for cloning *shd* because so far, the *shd* phenotype is characterised by a reduction in the number of iridophores alone (Kelsh et al. 1996). Because these pigment cells are not found in mammals there were no candidate genes for *shd.*

Positional cloning is initiated by identifying the broad location of the gene on a chromosome by linkage mapping. The sequential narrowing of the candidate genomic interval permits identification of the responsible gene whose function then can be studied by, for example, protein sequence comparison and gene expression studies.

Linkage maps are generated by breeding from a parent that carries two different alleles at two or more loci. The offspring are then tested for whether they receive either parental or recombinant allele combinations. Analyses of this type of data allow one to determine whether loci are linked to each other. For linked loci, their relative order and the distances that separate them can be estimated from the frequency of recombination events (Silver, 1995).

Simple sequence length polymorphisms (SSLPs) have become an important genetic tool in constructing linkage maps of vertebrates such as the mouse, rat and human (Rauch, 1997). These markers consist of two primers flanking a dinucleotide repeat, which can be highly variable in length and therefore used for genotyping individuals that come from a mapping-cross (Knapik et al., 1996). These genetic markers are highly polymorphic between the two backgrounds on which zebrafish mapping-crosses are built. Furthermore,they are codominant and abundant in zebrafish (Rauch, 1997). The SSLP map constructed for zebrafish, based on the Boston MGH cross, continues to be refined and updated and has been used as a framework in which to integrate higher resolution physical maps (Knapik et al., 1996; Knapik et al., 1998; Postlethwait et al., 1998; Gates et al., 1999; Geisler et al., 1999; Shimoda et al., 1999; Kelly et al., 2000; Woods et al., 2000; Hukriede et al., 2001; Iovine and Johnson, 2002; Stickney et al., 2002).

There are now seven mapping panels available, constructed by different labs in the zebrafish community: the Boston MGH Cross (MGH) has 3891 markers of which the great majority are SSLPs; the Gates et al (GAT) panel has only 464 markers; the Heat Shock (HS) panel has 7721 markers of which the majority are ESTs (Expressed Sequenced Tags); the Mother of Pearl (MOP) panel has 763 markers of which 257 are genes; the Goodfellow T51 panel has 10330 markers of which 4822 are ESTs and finally the LN54 panel with a total of 5000 markers of which 3417 are ESTs. The seventh mapping panel is called ZMAP and is an integrated map of the zebrafish genome which has a total of 23633

markers. This integrated map used the MGH SSLP based map as a framework for positioning all the coding markers. All these panels complement each other and are accessible for searches in the ZFin web page (http://zfin.org). These maps can be searched individually or in a consolidated version for various types of markers, genes, mutants and ESTs. They were generated by two different processes, based upon meiotic recombination (called meiotic maps) or by radiation hybrid mapping. Only the Goodfellow T51 and the Loeb/NIH/5000/4000 (LN54) panels fall in this last category.

Radiation hybrid mapping involves the fusion of irradiated zebrafish cells with a hamster cell line that is either hypoxanthine phosphoribosyl transferase (HPRT) or thymidine kinase (TK) deficient. The irradiation procedure randomly breaks the zebrafish genome, and the DNA fragments are rescued via the fusion process. Surviving cells are grown in a medium which selects for hybrids containing the zebrafish HPRT gene or TK gene. A dose of 3000 rads produces fragments in a size range required for map continuity and optimal resolution for cloning (Kwok et al., 1999). After extracting the DNA from the hybrid clones it is possible to use PCR for the ready mapping of any genomic markers by presence or absence. This includes genes, ESTs and other markers. The frequency of the radiation breakpoints can be translated in mapping distances that are more accurate than those provided by the meiotic maps. In this project we have used the MGH mapping panel which was, at the time, the diploid meiotic map with more SSLPs mapped. We did not use radiation hybrid mapping because as explained earlier there was not a candidate gene or EST for the *shd* locus.

The first stage in mapping a new locus involves genotyping of mutant embryos from a mapping-cross with a wide range of SSLP markers to find those which co-segregate with the mutant locus. This enables a rough positioning of the gene of interest on the zebrafish map and permits the identification of recombinant individuals in the mapping-panel. This information will be invaluable for streamlining the second stage, the genotyping of large

numbers of mutants with all closely linked SSLP markers. If these markers are still too distant (more than 1cM) from the mutant locus then other types of genetic markers such as amplified frequent length polymorphisms (AFLPs) (Ransom and Zon, 1999) have to be generated in the region flanking the locus. These are obtained by selective PCR amplification of restriction fragments from a total digest of genomic DNA. Because most AFLP fragments correspond to unique positions on the genome they can be exploited as landmarks in genetic and physical maps. Another alternative is to use Single-Strand Conformation Polymorphisms (SSCPs) (Fornzler et al., 1998; Kelly et al., 2000; Nechiporuk et al., 1999; Payne et al., 2001) which are polymorphisms found in non-coding regions of known genes and partially characterized cDNAs. A polymorphism frequency of approximately 50% has been demonstrated between the zebrafish strains used for genetic mapping studies (Fornzler et al., 1998).

The aim of the third stage of mapping is to narrow this genetic interval to a critical region, which is defined as a physical contig linking the two flanking markers that are no more than one recombinant apart from the mutant locus. On average, 1 centiMorgan (cM) in the zebrafish genome represents approximately 740 kb in a sex-averaged meiotic map (Postlethwait et al., 1999; Shimoda et al., 1999). So the closest linked marker, ideally less than 1cM away from the gene locus, is likely to be within reach of a relatively short genomic walk. The objective of this stage to bridge the gap between the nearest markers and the gene of interest. This involves a group of ever evolving techniques for genomic walking (Talbot and Schier, 1999) that detect overlapping sequences of individual clones and enable them to be positioned in the correct order. In particular, there are commercially available YAC (Yeast Artificial Chromosome), PAC (P1 artificial chromosome) and BAC (Bacterial Artificial Chromosome) genomic libraries (available from Research Genetics and Genome Systems) that can be screened by hybridisation or by PCR. These libraries contain average insert sizes of 240-480; 80-300 and 80-110 Kb of DNA, respectively (Amemiya et al., 1999). Therefore, if the distance to the closest marker is say, 1cM then a

walk of three or four YACs should cover the genetic interval containing the gene of interest.

A potential alternative method of delimiting the critical region containing the mutant locus involves microinjection rescue assays with genomic clones (Yan et al.,1998; Talbot and Schier, 1999). This approach may be very useful for easily scorable early phenotypes. *shd* has a phenotype that is readily scored at 72 hpf. At this stage, using a dissecting scope with incident lighting it is possible to identify the presence or absence of even single iridophores.

The aim of the work described in this chapter was to find the map position of both *colourless* (*cls*) and *shd* genes by linkage mapping of mutant embryos. The second objective was to limit the genomic region containing the *shd* locus to a minimum interval of no more than 1cM flanked by the two nearest markers.

2.2 Materials and methods

2.2.1 Mapping-crosses

Large sets of offspring were generated from fish that carry the *cls* and *shd* alleles. These mapping cross families were generated in the laboratory of Pascal Haffter (Max-Plank-Institut für Entwicklungsbiologie,Tübingen, Germany) by crossing strains having maternal and paternal chromosome sets from distinct genetic backgrounds (Tuebingen and Wik11) and which thus show the high degrees of polymorphism necessary for easy mapping (Rauch, 1997)(Figure 2.1). SSLP mapping on the F2 mutants and wild-type siblings from the mapping- crosses was performed using a semi-automated set-up during a visit to Tübingen in February 1999.

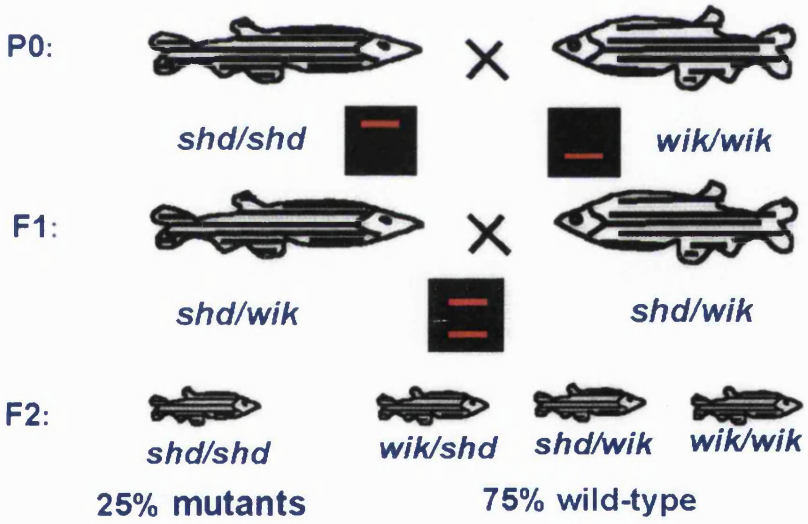


Figure 2.1. *shd* mapping-cross. Diagram showing how the mapping-cross was generated between the homozygous mutant *shd/shd* and the homozygous wild-type *wik11/wik11* (P0). These two strains can be easily recognised because they produce different size bands (red) when amplified with a polymorphic marker. The progeny of this cross (F1) are all heterozygotes and therefore have the same genotype carrying both mutant and wild-type bands. Crossing the F1 with each other generates the F2 generation. This generation contains 25% of homozygous mutants, which will be used for mapping, and 75% of wild-type siblings of which 50% have a heterozygous genotype and 25% are homozygous for the *wik11* allele. These siblings will also be used in the mapping process mainly for confirming the mapping of the mutants.

In order to extract the genomic DNA a total of 48 F2 homozygous mutants and 48 wild type siblings were collected from the respective map-crosses, identified and separated from their siblings, and kept in methanol at –70 °C in separate 1.5 ml eppendorf tubes. The embryos were then transferred to a 96 well plate (Thermoquick), allowing just one embryo in the bottom of each well. The DNA of the individual mutants and siblings was extracted with 2.25 ml of a mixture of 1 x TE (Sambrook, 1989) and 250 µl 17mg/ml Proteinase K (Sigma). Subsequently, 25 µl of this solution were transferred to each well with a multipipette (Finnpipette). The plate was covered with Hybaid film and heated in a Hybaid PCR machine at 55 °C for 250 minutes followed by 10 minutes at 75 °C. Finally, the genomic DNA extracted from the 48 mutant embryos was pooled in one tube and the same was done to the siblings' DNA.

2.2.2 Bulk segregant analysis

The PCR recipe used in the amplification of all SSLP markers was made up of 15µl of PCR mix, which was previously made in large quantities by mixing 6 ml of 10 x PCR buffer [10mM Tris pH 8.3 at 25°C (Sigma), 50 mM KCl (Sigma),1.5 mM $MgCl_2$ (Sigma) and 0.01% gelatine (Sigma)], which was added to 36 ml distilled water and 120 µl of each 100mM dNTP (GIBCO), 0.42µl of Taq polymerase 5 Units/µl (GIBCO) and 0.38µl of SSLP primers at 10µM (GIBCO). All these reagents were added to 5µl of 3 dpf embryonic genomic DNA.

The PCR program performed for amplification with SSLP markers consisted of a prior denaturation step of 2 minutes at 95 °C, followed by 34 cycles of 30 seconds at 94°C; 30 seconds at 60°C; 1 minute at 73 °C followed by 10 minutes final extension at 73°C. The amplified products were separated on 4-5 mm thick agarose gels. We used an overall 2% agarose gel made by mixing Metaphor (FMC) high quality agarose, with regular agarose

(Qualex). Gels were electrophoresed for 110 minutes at 100 Volts. We used 1x TBE buffer for making gels and running buffer and a type II loading dye (Sambrook, 1989). DNA was visualised with a UV transilluminator by prior staining of the gels with 20 µl,10mg/ml, of ethidium bromide (Sigma). Gels were recorded using an image analysis system (Sion Image) that could consistently record the band intensity produced by the pools of mutants and siblings. Any difference between the intensity of the two bands indicated a potential linkage to the SSLP marker in question. These markers were then tested on individual embryos for confirmation of linkage.

Extraction of DNA from individual embryos was done in the same way as described above except that DNA was not pooled in the final step. Instead, individual embryo DNA was kept separately in each well, diluted by adding 75 µl sterile distilled water. These plates were named template plates and used for the amplification of one marker at a time. The amount of DNA present in each well allowed for up to 20 PCR reactions, *i.e.* for the genotyping of the same individuals with up to 20 different markers. The PCR products were obtained and visualised in the same way as mentioned above. To resolve some of the polymorphisms pure 2% Metaphore agarose (FMC) was used (see Figure 2.2 for example of how a closely linked and distant marker looked on a gel).

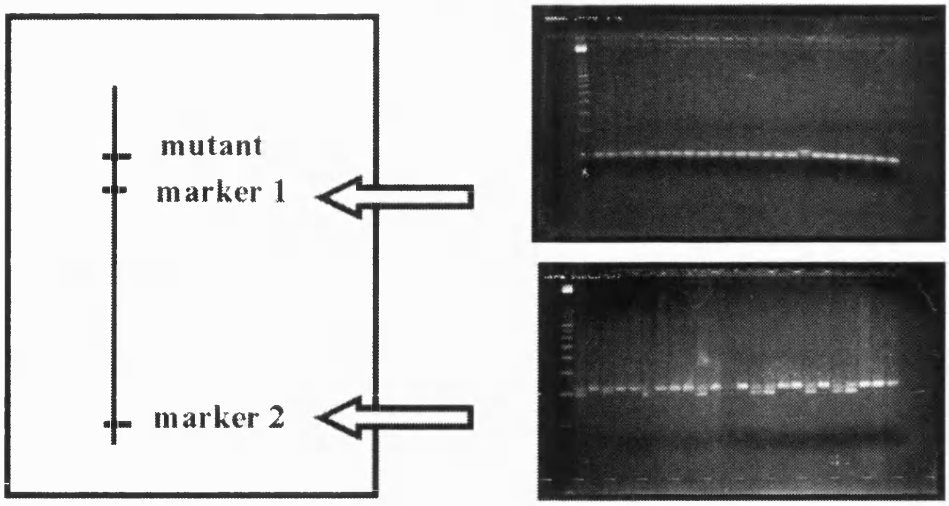


Figure 2.2. Confirmation of linkage by genotyping individual mutant embryos with different SSLP markers. Example of a closely linked marker 1 with the mutant locus and a more distant marker 2. The closer the marker is to the mutant locus the fewer recombinant embryos will be present. A recombinant mutant embryo will be heterozygous for the mutant and wild-type alleles because one or more cross-overs must have occurred between the two loci. So, the same mutant embryos when genotyped with marker 1 produce a lower size mutant band and a higher size wild-type band and when genotyped with marker 2 generate the opposite situation but this time presenting many more heterozygotes due to the increased number of recombinants.

Linkage of markers with the mutation in question was evaluated using a computer script written in Filemaker Pro 4.1 by Dr. Robert Geisler. Gels were scored by assigning numbers to the different bands. Embryos homozygous for the upper band were scored as 1, for the lower band as 2 and heterozygotes were scored as 3.The linkages given in cM and the respective LOD scores for each marker were then calculated by this script allowing us to position the markers in an ordered array in relation to the mutant locus. The LOD score was calculated using the following equation (Geisler, R. pers. Comm.)

$$LOD = lg\ [exp\ (2*n*(ln\ (1-\theta)*(1-\theta) + ln(\theta*\theta - ln\ 0.5)))]$$

Where:

n = total number of diploid embryos

θ = recombination fraction

2.2.3 Fine mapping of the *shd* locus

To fine map the *shd* locus large numbers of F2 embryos (~1000) were genotyped using the closest markers to the mutant locus. This enables more accurate determination of the distance of the markers to the mutant locus. In order to do this we brought F1 embryos of the *shd* map-cross from Tuebingen to Bath. These embryos were raised in our fish facility under intensive feeding conditions (4 feedings per day) using a variety of flake food (Tetramine) growth food (Tetramine) and hatched artemia. F1 carriers were identified and pairs kept separately. F2 progeny from individual F1 pairs were collected weekly, DNA was extracted and each individual was genotyped with SSLP markers using the PCR conditions described above.

To enable mapping of all new SSLP markers in the region of *shd* locus, DNA from the parental (P_0) generation of the map-cross was tested for polymorphisms. DNA was generated from tail fin clips and all polymorphic markers were then genotyped on the F2 embryos and became part of our mapping panel. Tail fin clips were taken from adult fish after anaesthesia in 0.04% MESAB using a clean scalpel. Fish were returned immediately to fresh water. After recovering they were returned to their original tank. Each tail was placed in eppendorf tubes in 100µl lysis buffer (50mM Tris pH 8.3, 100mM NaCl, 2% SDS, 5mM EDTA, 200µg/ml proteinase K (Sigma)), and were incubated at 55°C overnight for digestion. The next day the DNA was cleaned by adding 5 to 6 ml of phenol / chloroform / isoamylalcoholl 25:24:1 mixed thoroughly, spun for 35 minutes at 5000 rpm, and then the supernatant was removed with a truncated blue tip. The supernatant was transferred to a clean tube and the DNA precipitated by adding 0.3 M NaCl , 2 volumes of cold 100% ethanol and spinning for 15 minutes at 5000 rpm. The DNA pellet was washed with 70% ethanol and left to air dry for 30 minutes. The pellet was resuspended in TE and the DNA was stored at 4°C ready to be used.

2.3 Results

2.3.1 Mapping the *cls* and *shd loci*

Both *cls* and *shd* loci were positioned on the zebrafish linkage map (MGH mapping panel). Mapping cross families for these two loci had been generated by crossing the mutants, which were firstly generated on a Tuebingen background (TU) (Haffter et al., 1996a), to the wild-type strain WIK11. This strain had been previously tested for polymorphisms with a large number of SSLPs by (Rauch, 1997) and proven to be the most suitable for mapping. Different sets of SSLPs were amplified from pools of genomic DNA from mutants and wild-types. After using several sets of markers (each set has 48 different markers) the whole genome (2635 cM) has been scanned to a resolution of less than 2 cM thus allowing linked markers to be identified.

As explained above, genotyping the homozygous F2 embryos from the respective map-crosses permitted to find linkage to a few markers and therefore map the two loci to their respective linkage groups. In this way, *cls* was placed on linkage group 3 (not shown here; see Dutton et al., 2001) and *shd* on linkage group 17 (Figure 2.3). In 274 meioses the closest markers flanking *cls* were z872, 2.4 cM away and z13387, 1.5 cM away. The closest flanking markers to the *shd* locus in 540 meioses were z10985, 0 cM distant and z11341, 2.0 cM away. The purpose of mapping these two mutations was very different. While *cls* was being cloned by a colleague, Angela Pauliny, using a candidate approach, *shd* was going to be positionally cloned. Thus, we needed only a rough map position for the *cls* locus, so that my colleague could then map her candidate gene onto a radiation

hybrid map and see if the two map positions coincided. On the other hand, we needed a very reliable and precise map position for *shd* for the positioning cloning to be feasible.

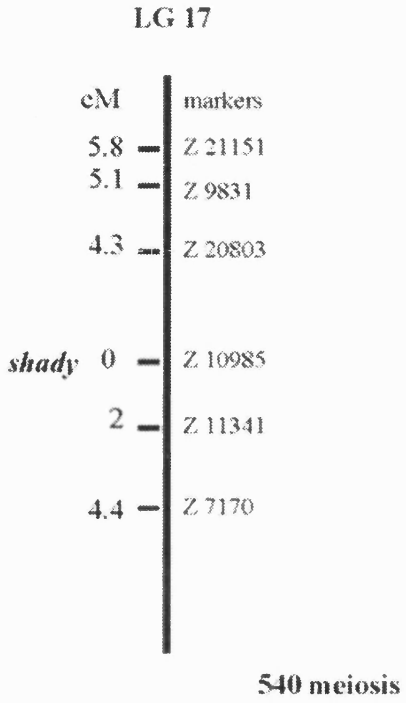


Figure 2.3. Diagram showing the results of the first stage of mapping *shd*. The mutant locus was placed on linkage group 17 between the markers z20803 and z11341 comprising an interval of 6.3 cM. At this stage the marker z10985 did not show any recombinant with the mutant locus and therefore it was impossible to know on which side of the *shd* locus it lay. The centromere was located further south of this set of markers.

2.3.2 Fine mapping of the *shd* locus

After waiting for the *shd* map-cross F1 fish to mature (brought as embryos from Tuebingen in February 1999), F1 identified carriers were crossed and large numbers of F2 mutant embryos were collected for genotyping with the best SSLP markers found in the first mapping stage. A total of 1000 mutant embryos were therefore genotyped with five different markers. We will present two examples to illustrate the process of mapping used.

The marker z10985, the closest marker, had a typical size of 99bp on a Tuebingen background and around 111bp on a Wik11 background (see Figure 2.4 for an example of a gel produced during mapping with this marker).

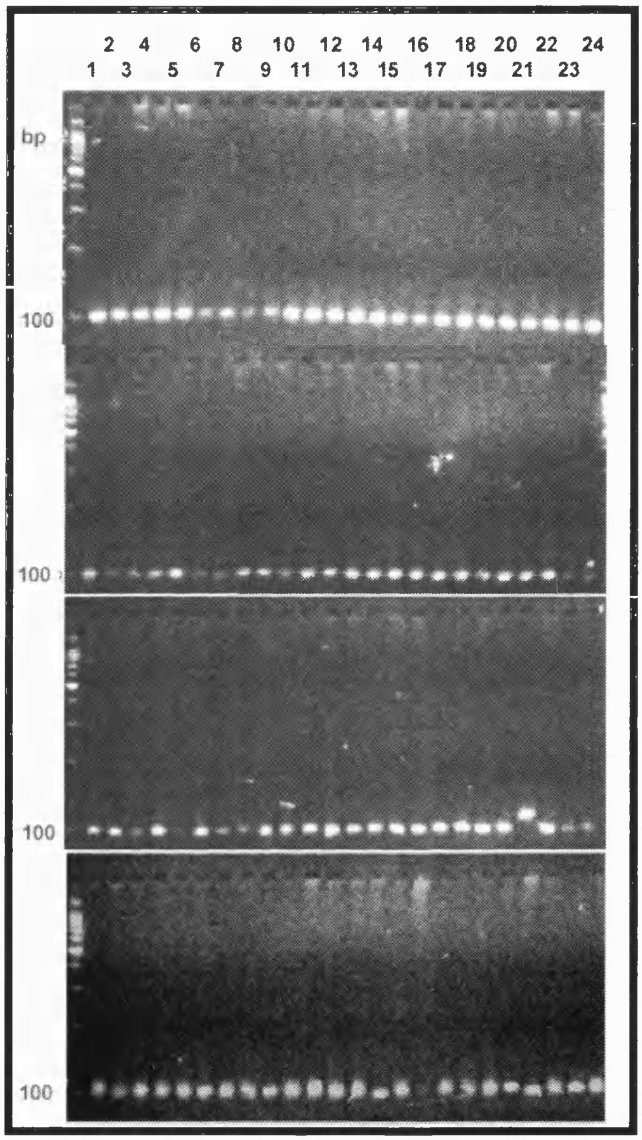


Figure 2.4. Example of a typical gel showing the mapping of z10985 on 96 individual embryos. The first three gels are the result of amplifying z10985 on *shd* homozygous mutant embryos. The last gel was always used for amplifying the same marker on sibling embryos from the same cross. All the mutant embryos show the same 99bp band for the marker z10985. There is one exception on the third gel showing a clearly mis-sorted embryo amplifying the larger band typical from a WIK11 background. This is a sibling embryo homozygous for the WIK11 band that was picked by mistake and included among the mutants. In the last gel it is possible to see several of these embryos homozygous for the Wik11 allele (e.g.numbers 20, 22 and 23 from left to right). The majority of the siblings are heterozygotes for both alleles and there are two mis-sorted mutant embryos homozygous for the lower band (numbers 14 and 21).

Several gels like this one were obtained for this marker and only one recombinant was found in the total of 1000 mutant embryos genotyped. Therefore the marker z10985 showed only one recombination event with the *shd* locus indicating a distance of 0.1 cM, which according to Shimoda et al. (1999) translates to around 74kb. By comparison, the next closest markers are significantly more distant. Thus, z11341 showed 64 recombination events on the other side of the *shd* locus and the marker z20803 scored 70 recombinants on the same side of the *shd* locus as z10985. Figure 2.5 shows a typical example of a mapping gel for one of these markers. An explanation of how we can tell the orientation of each marker in relation to the *shd* locus is given further below.

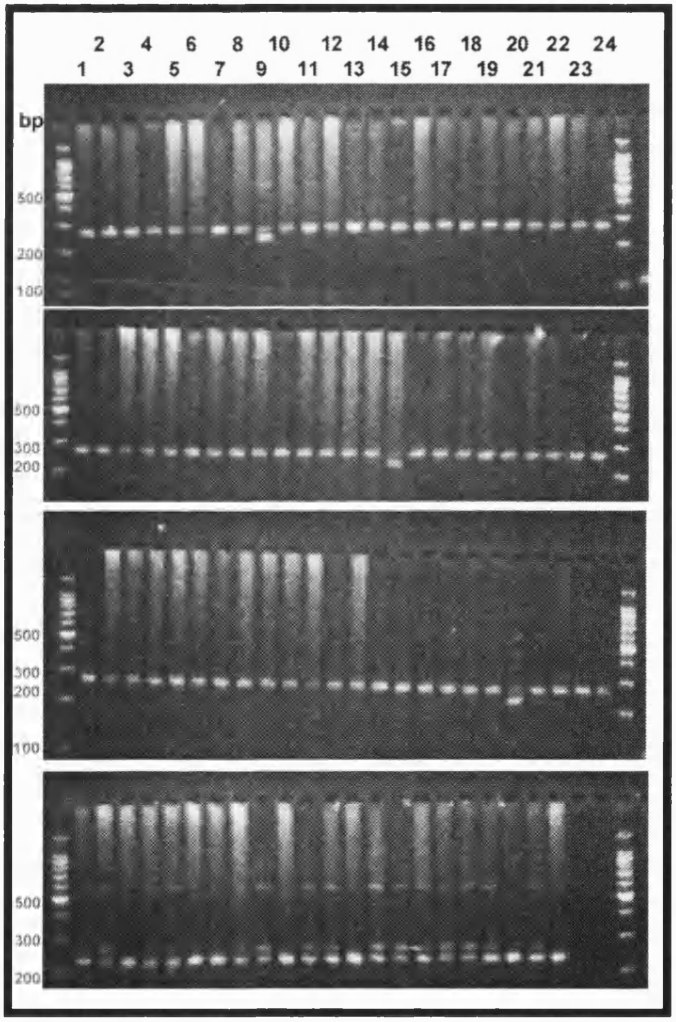


Figure 2.5. Example of a typical gel showing the mapping of z11341 on 96 individual embryos. The first three gels are the result of amplifying z11341 on *shd* $^{-/-}$ mutant embryos. The last gel shows mapping of the same marker on sibling embryos from the same cross. Most mutant

embryos show the same 250bp band. However, there are also three recombinant embryos that show a lower band around 220bp and also a much higher and weaker band around 600bp. In the siblings gel it is possible to see many of these genotypes heterozygous for the Tuebingen and Wik11 alleles due to presence of wild-type heterozygous embryos which have the same genotype. The third genotype seen in the siblings gel corresponds to the homozygotes for the Wik11 allele, a single band around 220bp.

These two markers indicated an average distance to the mutant locus of around 7cM, i.e. approximately 5Mb. This distance was too great to consider any type of genomic walk. The remaining markers we had been consistently testing, z7170 and z9831, were even further away (see Figure 2.6 for an example of a gel for the mapping of z9831).

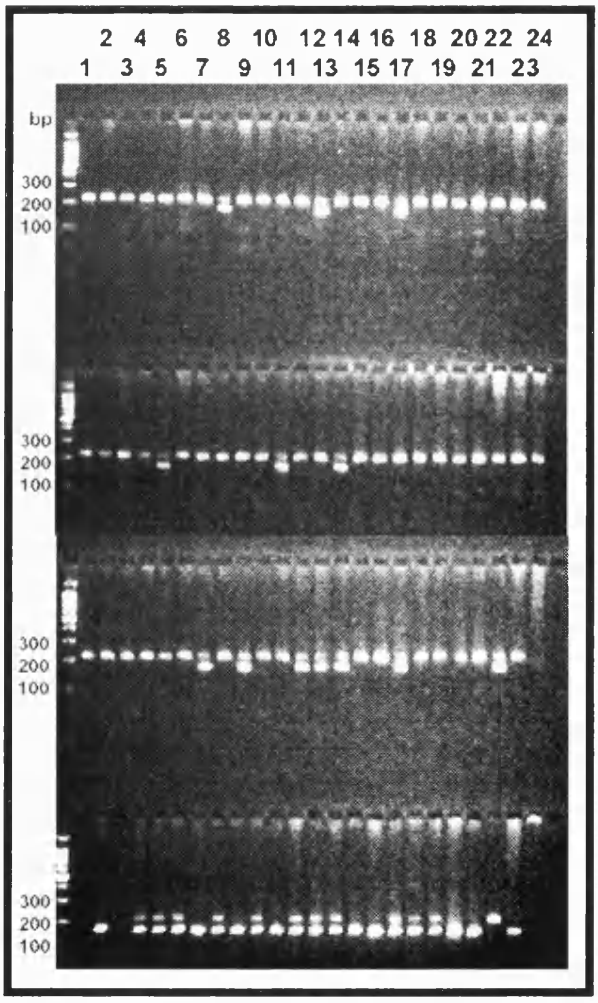


Figure 2.6. Example of a typical gel showing the mapping of z9831 on 96 individual embryos. The first three gels are the result of amplifying z9831 on shady homozygous mutant embryos. The last gel tests the same marker on sibling embryos from the same cross. Most mutant embryos show the same 210bp band typical for the Tuebingen background. However, there are 13 mutants showing both this band and the Wik11 band at around 190bp. These recombinants are *shd* $^{-/-}$ embryos that are heterozygous for this SSLP. Thus, their genotype is the same as that of the heterozygous wild-type siblings. The third genotype presented in the

siblings gel corresponds to the homozygotes for the Wik11 allele, a single band around 190bp. It is also clear in the siblings gel that there is one mis-sorted embryo at position 22 from left to right, this is a mutant embryo homozygous for the Tuebingen allele.

Testing of these more distant, as well as the closest, markers for all 1000 embryos was important as a control for mis-sorted embryos. A mis-sorted embryo will typically be homozygous for the WIK11 allele, and thus will be easily detected as it will show a single different band on the gel. More confusingly, mis-sorted wild-type embryos can be heterozygotes for the WIK11 and TU alleles, and thus can be easily confused with a recombinant embryo (but homozygous mutant). The only way to detect such mis-sorted embryo is by using several markers mapping to the region of the mutant locus. Amplification of different markers mapping to the same side of the mutant locus should identify the same embryos as recombinants. The number of recombinants is expected to be decreased for markers mapping closer to the mutant locus. Heterozygote embryos that do not comply with this pattern can be unambiguously identified as mis-sorted wild-type embryos, not recombinant *shd* mutants.

At this stage the DNA from each of the recombinant embryos was physically removed from the original template plates and put into separate new plates. One for recombinants distal to *shd* and one for recombinants proximal to *shd*. These new plates constituted the definitive recombinant mapping panels on each side of the *shd* locus. Instead of genotyping every new polymorphic marker on the same 1000 mutant embryos, they would only be tested on the recombinant mapping panels. Any new marker that showed fewer recombinants than z11341 (64) or z20803 (70) would be closer to the *shd* locus. We tested several further markers and three new markers mapped closer than the previous flanking ones. Thus, z15715 mapped on the same side as z11341 but showed only 26 recombinants and z9818 and z21435 mapped to the same side as z20803 (the same side as z10985) but scored 20 and 40 recombinants, respectively. An example of the mapping of one of these markers (z9818) using the recombinant panels is presented in Figure 2.7.

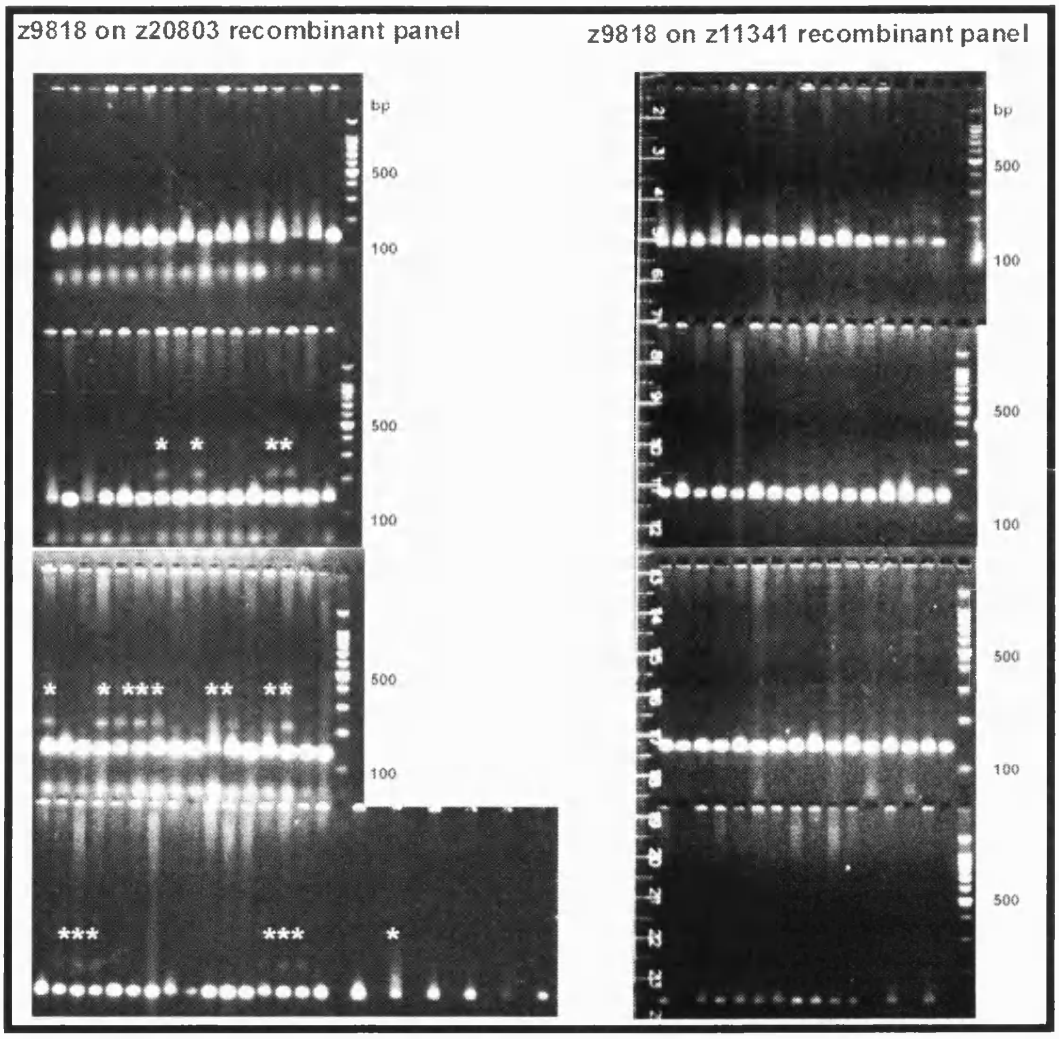


Figure 2.7. Mapping of the marker z9818 on the two recombinant mutant panels showed 20 embryos recombinant for this marker (*). As expected, no recombinants were found on the z11341 panel, meaning that the z9818 is on the same side as z20803 in relation to the *shd* locus. Z9818 shows a mutant band around 150bp and a Wik11 band around 210bp

Figure 2.8 summarises the mapping data for these markers. The marker z10985 was still much closer than any other marker tested. At that time there were no other polymorphic markers available in the *shd* genomic region.

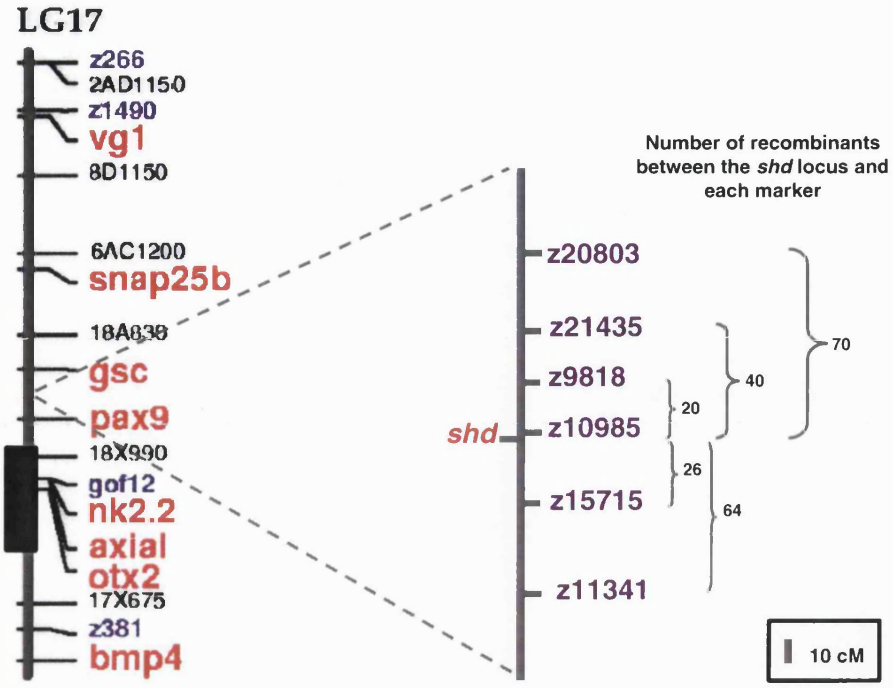


Figure 2.8. Diagram showing the position of *shd* in relation to other genes and markers mapped to linkage group 17. The insert shows the *shd* region to indicate the relative positions of the zmarkers that were mapped on 1000 individual mutant embryos. The number of recombinants between the *shd* locus and each marker is indicated. 1 recombinant found in 1000 mutant embryos translates to around 74 kb in a sex averaged map (diagram adapted from Postlethwait et al., 1998)

At this stage the *shd* locus was flanked by markers that limited a genomic interval of about 2Mb. This distance was still too great to perform a bi-directional genomic walk. We thus faced a choice between a) generating new markers and looking for closely linked markers on the z11341 side or b) walking from z10985. The distance from z10985 alone was within the reach of a very short genomic walk, i.e. 74 kb which was within the standard size of one PAC clone. Consequently, the plan was to screen a genomic library with the marker z10985, isolate a number of PACs and analyse these by a combined approach including restriction analysis, sequencing and microinjection. We hypothesised that one of the identified PACs would probably contain the complete *shd* gene.

2. 4 Discussion

We have mapped *shd* on linkage group 17. The *shd* locus is flanked by two markers: a closely linked one and a more distant one. Because between this more distant marker and the *shd* locus we still found 26 recombinants we decided to start a unidirectional walk from the other flanking marker with which we had found only one recombinant mutant embryo.

Positioning *shd* on zebrafish LG 17 raised some interesting questions with regard to conserved syntenies. Previous analysis revealed that a group of genes that is syntenic (on a single chromosome) in humans tend to have orthologues that are syntenic in zebrafish (Postlethwait et al., 1998). The reason it is preferred to compare zebrafish genomics with human genomics arises from the fact that in general the zebrafish genome seems to correspond more closely to the human genome than the mouse genome. This reflects the rapid chromosome evolution among rodents and the rather slow divergence of karyotypes in the primate lineage (Postlethwait et al., 1999).

Woods et al. (2000) showed that zebrafish genes on linkage group 17 have their orthologues in clustered regions of human chromosomes 2p; 10q; 14q; 20p. Of these human clusters the most numerous are chromosome 14q with 10 genes and chromosome 20p with 5 genes. These syntenic genes include two genes, *pax9* and *gsc*, that flank the *shd* locus (see Figure 2.8). Both these genes are in the 14q human cluster. Thus, we can speculate that perhaps that entire zebrafish genomic region, including the *shd* locus, may be syntenic to human chromosome 14q.

However, at present time, with more genes being positioned on the zebrafish maps everyday, further investigation showed that zebrafish genes *otx1* and *id2* (both on LG 17

and even closer to the *shd* locus) have orthologues on human chromosome 2p13 and 2p25, respectively (http://zfin.org/cgi-bin/view_zmapplet.cgi.). Furthermore, *Sox11a* and *gdf7* were also mapped on this region (Davidson et al., 1999; de Martino et al., 2000) and their human orthologues are found on 2p25 and 2p24-23, respectively (http://www.ncbi.nlm.nih.gov). In this way, there are at least four genes confirming the synteny between this confined region of zebrafish linkage group 17 and the human chromosome 2p13 - 2p25 interval. It is therefore possible to speculate that if there is such a human orthologue of the *shd* gene it will probably lie in this region of the human genome. However, at the stage that this mapping phase was completed there were very few genes positioned on linkage group 17 and it was impossible to draw such speculation. For this reason, a candidate approach by looking at all these human genes and evaluating their potential candidacy from their function in development or expression patterns was not undertaken. Instead, it was decided to continue with the positional cloning by narrowing the interval within which the *shd* gene was expected to be.

Chapter 3

Isolation of the genomic region containing the *shady* locus

3.1 Introduction

3.1.1 Genomic libraries

In order to study fully the function of genes identified by the mutagenesis screens it is necessary to positionally clone these genes (Collins, 1992). This task is not trivial and can be labour-intensive, requiring an arsenal of both genetic (e.g. different types of genetic markers) and genomic tools (e.g. somatic cell/radiation hybrids, genomic libraries). First, we will give a brief description of the three types of zebrafish genomic libraries available at this stage of the project: the YAC system constructed in yeast artificial chromosome, the BAC system in bacterial artificial chromosome, and the PAC system constructed in P1 derived artificial chromosome vectors. All these cloning systems have been used successfully in the human and mouse genome projects and they are expected to have the same impact in the zebrafish positional cloning projects (Amemiya et al., 1999).

The YAC cloning system (Nelson, 1990) uses two vector arms, both containing yeast selectable markers and telomere sequences. One of them contains a centromere and a yeast autonomously replicating sequence (ARS). The (dephosphorylated) vector arms are ligated to the size-selected genomic DNA to be cloned (usually Eco RI partial digests) and subsequently transformed into yeast cells which have had their cell walls enzymatically removed (spheroblasts). To ensure that only clones containing both vector arms are propagated, the transformed spheroblasts are plated in top agar prepared with selective media. Clones with inserts are further detected by a red-white selection. Transformants comprise features necessary for mitotic propagation in the yeast cells: namely two telomeres, an ARS, a centromere and presumably a genomic insert. This system is highly advantageous in that it is

capable of propagating extremely large (megabase-sized) fragments. However, the disadvantages of this system are many and include the relatively high rate of insert chimerism, inherent instability, slow growth rate and overall difficulty in handling. As YACs are linear chromosomes, they can only be isolated from the endogenous yeast chromosome by separation via pulsed field gel electrophoresis which is a relatively tedious procedure.

The BAC system (Shizuya et al., 1992) is bacterial plasmid-based. The system operates under the principle that the *Escherichia coli* F-factor plasmid replicon is capable of replicating DNA molecules as large as the entire *E. coli* chromosome. The vectors used for BAC cloning have been stripped of all accessory components and are essentially an F-factor replicon with a chloramphenicol resistance gene (Shizuya et al., 1992). The latest BAC vectors all possess a Lac-Z blue-white colour selection system. As with the YAC system, partially digested genomic DNA (usually Hind III) is extensively size-selected and ligated to dephosphorylated BAC vector (7.5 Kb). Ligation products are then electrotransformed into a *E.coli* host strain that possesses a cell wall defect (DH10B; Life technologies) and recombinants isolated by plating on selective media. The BAC system can only propagate inserts up to 300 Kb. However, this system is highly advantageous in that BACs are essentially large plasmids and can be handled and characterised using simple modifications of routine methods used in all molecular genetics laboratories. BACs are also relatively stable and do not exhibit the high levels of chimaerism observed in YACs (Amemiya et al., 1999).

The PAC cloning system (Ioannou et al., 1994; Ota and Amemiya, 1996) is very similar to the BAC system, except that is uses a (single copy) P1 bacteriophage replicon and a Kanamycin resistance gene. As with the BAC system, size-selected partial digests (usually with MboI or Sau3A) are ligated to a desphosphorylated vector and electrotransformed into DH10B cells. The PAC vector (~16Kb) additionally contains a SacBII selection system against noninsert-

containing clones, whereby only recombinant clones grow vigorously in the presence of 5% sucrose. The PAC system can accommodate inserts of up to 300Kb and exhibits essentially all the advantages of BACs. In addition, it contains an inducible P1 lytic replicon that permits multicopy propagation of PAC plasmids when the clones are grown in the presence of 1mM IPTG. The disadvantage of the PAC system with respect to the BAC system is the larger vector size.

To date, there are two zebrafish YAC libraries, one BAC library, and one PAC library. All these libraries are available (Genome Systems and Research Genetics). All the libraries encompass from four to five haploid genome equivalents. The average insert sizes of the four libraries range from 470 Kb to 90 Kb and there is roughly a 20x genomic coverage (Amemiya et al., 1999). The three vector systems are complementary in a sense that different genomic sequences may be under or over-represented in each system.

The first aim of the work described in this Chapter was to screen the zebrafish PAC library with the marker z10985 in order to isolate a maximum number of PACs that would contain, not only this z marker, but also parts of the *shd* gene or even the complete *shd* gene. This marker was estimated to be around 74 Kb distant from the *shd* locus and therefore the choice of a genomic insert in the range of the PAC was the best suited. Three PACs containing the marker were isolated.

3.1.2 Identification of clones containing the mutant gene by genomic DNA microinjection

Taking a phenotypic rescue approach, at least at an early stage of a positional cloning project, is not common. In fact, so far only one project among the published examples of zebrafish positional cloning has followed this procedure (Guo et al., 2000). At the time this assay was initiated in the current work, there were no published examples of genomic rescue as a means to positional clone a gene. Nevertheless, it was known that *floating head (flh)* mutants could be partially rescued by injection of BAC clones containing the wild-type gene (Yan et al., 1998). Phenotypic rescue was only partial, presumably reflecting the mosaic inheritance and expression of the injected transgene. This mosaicism had been documented before for other clones injected into one-cell stage embryos (Westerfield et al., 1992). This fact meant that for some embryos, with very subtle phenotypes, affecting only a small number of cells, the partial rescue might not be detectable. Another unknown factor, at this stage was how long transient rescue would be effective. The *shd* phenotype is not seen reliably until 72 hpf and it was unclear whether iridophores could be rescued in a transient assay, nor how much genomic DNA it would be necessary to inject to have such a late effect.

These potential problems were perhaps the reason why this assay was not very common. Instead, the standard procedure was to create more markers in the ends of identified genomic clones for continuing the genomic walk (Talbot and Schier, 1999). This demanded that the ends of the genomic clones were sequenced and that specific primers were synthesised and then tested for overlap and suitability to serve as Sequenced-Tagged Sites (STSs). This also meant that every new STS had to be mapped in the recombinants mapping panel for confirmation of linkage and assessment of the progress of the walk. Polymorphisms in STSs

may be detected with restriction enzymes or by single-strand conformational analysis (Talbot and Schier, 1999).

In summary, the alternative to phenotypic rescue with genomic microinjection is laborious. So, the second aim of this Chapter was to optimise the PAC injections to enable phenotypic rescue of iridophores by trying a number of concentrations of genomic DNA and carefully assessing the injected embryos after 72 hpf.

3.1.3 Assembly of a genomic region

Once there are identified genomic clones it is essential to analyse them by pulsed field electrophoresis to know their size and to assess the extent of overlapping regions (Amemiya et al., 1999; Talbot and Schier, 1999). It is then of major advantage to construct a map with these results in order to visualise the genomic contig. In this map of the assembled genomic region, it is possible to identify important markers found previously and PAC end-clones by hybridisation techniques.

Perhaps the most powerful outcome of a correct assembly of the genomic region containing the locus of interest and identification of a PAC that produced transient rescue after microinjection was that at this stage it was possible to pinpoint exactly the region containing the gene and its regulatory elements required for the rescue. It meant that, in theory, it would be possible to sequence only the fragment containing the gene, instead of a large genomic stretch. Additionally, it would also be possible to isolate a coding fragment of the gene by direct cDNA selection or exon trapping (Amemiya et al., 1999).

In this way, the third objective of this Chapter was to assemble the genomic region where *shd* lies in and combine that information with the other results obtained so as to identify the critical region containing the gene.

3.1.4 Identification of candidate loci within a genomic region

There are several approaches to tackle the last stage of a positional cloning project : the identification of candidate genes. cDNA libraries can be screened with the large genomic clones or a fragment from them that is known to contain the gene of interest. However, this approach presents some problems like the repetitive sequences present in untranslated sequences of some cDNAs (Talbot and Schier, 1999) and even after resolving this type of problem it usually results in identifying several false positives (e.g. Brownlie et al., 1998). These will all have to be tested, a process which can be very time consuming. Another problem with working with cDNA libraries is that it requires prior knowledge of when the gene is active. Nevertheless this approach has been followed earlier by (Brownlie et al., 1998) to clone the *sauternes* gene and by (Zhang et al., 1998b) to clone zebrafish *one-eyed pinhead.* More recently, the zebrafish mutation *weissherbst* (*weh*) was positionally cloned by screening cDNA libraries with large genomic inserts. *weh* was identified as *ferroportin1,* a conserved vertebrate iron exporter (Donovan et al., 2000). The positional cloning of the mutant *miles apart* (*mil*) as a sphingosine-1-phosphate receptor presents another example of successful screening of cDNA libraries with PAC inserts (Kupperman et al., 2000).

Another, less problematic, approach is to sequence the whole critical region if this is small enough. With the ongoing zebrafish genome sequencing project this is made ever easier

since the Sanger Centre has prioritised the sequencing of known contigs. This proved feasible for the *shd* region, thus enabling us to analyse the full genome sequence of this area.

Whether we have several candidate cDNAs or a large sequenced region containing the gene of interest, it still requires the testing of potential candidates. In the case of candidate cDNAs it is possible to perform whole mount in situ hybridisation, a routine method in zebrafish. This is likely to be informative unless the gene in question is ubiquitously expressed. On the other hand, having a large genomic sequence to analyse may be arduous. Fortunately, there are sophisticated computer programmes made available by the UK Human Genome Mapping Project Resource Centre (UK HGMP-RC) that aim to identify potential genes in large genomic sequences (a more detailed explanation on these programmes is given in the Results section of this Chapter).

Candidate genes identified in this 'virtual' way have to be confirmed experimentally. In other words, the most promising ones have to be cloned and tested by whole mount in situ hybridisation and then by either rescuing the mutant phenotype by mRNA or cDNA injection or by phenocopying the mutant phenotype by injecting morpholinos. Almost all published positional cloning projects have shown mutant rescue by cDNA injection (Brownlie et al., 1998; Donovan et al., 2000; Guo et al., 2000; Kupperman et al., 2000; Shafizadeh et al., 2002; Zhang et al., 1998a) and few by using morpholino technology (Parsons et al., 2002). The one exception to this pattern is *gata1* for which both mRNA and cDNA failed to rescue the phenotype of the *vlad tepes* mutant. In this particular case injection of a BAC containing the *gata1* gene and assessment of rescue identified *gata1*as the mutant locus (Lyons et al., 2002). Ultimately, in the molecular identification of any mutated gene, it will be always necessary to discover the molecular lesions present in the mutant alleles of the gene (Talbot and Schier, 1999).

Thus the last aim of this Chapter was to 'virtually' identify one or more candidate genes for the *shd* locus. The combination of all the results from this Chapter helped find the most plausible candidate in preparation for definitive testing of that candidate (see Chapter 4).

3.2 Materials and methods

3.2.1 Screening P1 artificial chromosome genomic libraries

3.2.1.1 DNA-pool-screening procedure

All RZPD libraries can be screened by PCR in a convenient two step procedure. For each step sufficient DNA was received for at least 25 rounds of screening. The purified DNA was prepared by RZPD using an Automated Nucleic Acid Isolation System and was submitted lyophilised in a 96 well format microtiter plate. Initially, the primary pools from the zebrafish PAC library 706 created by Chris Amemiya from erythocytes (Amemiya and Zon, 1999) were analysed. DNA was resuspended in MilliQ water and 5µl of purified DNA was obtained from all clones from 33 microtiter plates, which equals 33x96 clones. The same PCR procedure described in Chapter 2 (section 2.2.2) for the mapping of marker z10985 was applied to all the clones. Products were run on 1% agarose gels.

Secondary pools consisted of individually pooled microtiter plates from the positive primary pool procedure. The secondary pool screening (also performed using the same PCR amplification) had to produce at least three positive signals to identify: the positive plate, the positive row and the positive column. The combined information of the positive signals identifies the unique clone. Subsequently, the identified clones can be requested from the RZPD Resource Centre.

3.2.2 Rescuing the *shady* phenotype by genomic DNA microinjection

3.2.2.1 Isolation of PAC DNA

The identified clones were ordered from the RZPD Resource Centre and were sent as stabs. Glycerol stocks were made as per Sambrook et al., (1989) and kept at –80°C. Colonies were grown overnight in the presence of 20mg/ml of Kanamycin. In the next day individual colonies were picked and cultured overnight in 300 ml of LB with 20 mg/ml of kanamycin. The QIAGEN large-construct Kit (QIAGEN) was used for the purification of PAC DNA according to the manufactors' instructions. This kit is designed to isolate DNA from large constructs such as BACs, PACs, and cosmids of up to 250 kb in size. The kit uses a rapid purification procedure based on the selectivity of a patented QUIAGEN Anion-Exchange Resin. A digestion step with ATP-Dependent Exonuclease enables efficient removal of genomic DNA contamination to yield pure genomic DNA free of vector DNA. However, because these large constructs are typically present at very low-copy numbers yields can sometimes be low using this column based method.

To increase yields the isolation of PAC DNA from 100 ml cultures employing the Plasmid-Safe ATP-dependent Dnase enzyme (Epicentre Technologies) was also used. This enzyme digests away all DNAs except those that are circularised. The resultant DNA is sufficient for subcloning and restriction analysis, and can be scaled up if necessary. The full protocol is available at http://med-humgen14.bu.edu/vectors/PACisol-100ml.h. The procedure involved inoculating 100 ml of LB-Kanamycin (25 μg/ml) with a 1 ml pre-inoculum of the desired PAC clone. A 500 ml erlenmeyer flask was used to grow the culture at 37°C overnight at 300 rpm. Next day the culture was spun at 5000 rpm for 5 minutes at 4°C and the supernatant

discarded completely. The material was resuspended with 8 ml of P1 solution (Sambrook et al. (1989) and transferred to a 50 ml SA600 (SS34) screw-top polypropylene tube and vortexed briefly to completely suspend the bacteria. 8 ml of the P2 solution (Sambrook et al. (1989) was added and gently mixed by swirling the tube. The solution was allowed to sit on the bench for 5 minutes, swirling occasionally, before adding the P3 solution and inverting the tube a few times to mix. The tube was placed on ice for 10 minutes and then was spun in a SA600 rotor for 15 minutes at 12000 rpm at 4°C, supernatant discarded and all traces of isopropanol removed. The pellet was washed with 5 ml of 70% ethanol by inverting the tube several times, before spinning again at 12000 rpm for 5 minutes and discarding the supernatant. The pellet was air-dried briefly until it became translucent and 500 μl of Plasmid-safe buffer (Epicentre technologies) was added and swirled several times to resuspend the DNA. Then, 5μl of 100 nM of ATP, 5μl of plasmid-safe ATP-dependent Dnase and 1μl of Rnase (10 mg/ml) were added and the tube was placed in a 37°C water bath for 3 hours with occasional swirling. The DNA was then transferred to a 2 ml eppendorf tube and phenol-extracted followed by two phenol-chloroform extractions. DNA was precipitated by adding 50μl of 3M NaOAc and 1.2 ml of absolute ETOH. The tube was placed on ice for 10 minutes, then spun in a microcentrifuge for another 10 minutes. The supernatant was completely removed and 1.8 ml of 70% ETOH was added to wash the pellet by inverting the tube several times and then spinning it for 1 minute and removing all traces of ethanol with a pipette. In the end the pellet was resuspended in 100 ml TE by leaving the DNA at room temperature over the weekend.

Prior to injecting the PAC DNA into zebrafish embryos another cleaning step was introduced by using the Microcon-100 Centrifugal Filter Devices (Millipore) according to the manufactors instructions. This step was also applied to concentrate the DNA whenever necessary.

3.2.2.2 Microinjections with PAC DNA

Fish that were identified as carriers for the strong *shd* $^{ty82/ty82}$ mutation were inter-crossed the day before injection was performed. Eggs laid on the morning of the injections were carefully cleaned with embryo medium (5mM NaCl, 0.17mM KCl, 0.33mM $CaCl_2$, 0.33mM $MgSO_4$, 10^{-5}% methylene blue to suppress fungal growth) and observed under a dissecting microscope with the purpose of checking that there were not any abnormalities detectable at this early stage. Eggs were then placed in the groove of an injecting petri dish coated in 2% agarose according to (Westerfield, 1995). Eggs were trapped and aligned in the groove to hold them stable during the injection procedure. Each groove takes about 50 eggs from which only 25% were expected to be mutants. The aim was to inject around 400 embryos in order to sample 100 mutants.

The injections were done using a Nanoject II injector (Drummond) and the needles were made of borosilicate glass capillaries GC120F-10 (Clark Electromedical Instruments) that were pulled in a Sutter instruments Co. Model P97 puller. The concentration of the PAC DNA was checked using a spectrophotometer (Beckman DU530) and diluted in MQ water and 1% phenol red to three different final concentrations: 68 pg/nl, 25 pg/nl and 50 pg /nl. The injector was set to release approximately 10 nl of liquid; thus each embryo would have received 680, 250 and 500 pg of DNA, respectively. The injection was done into the yolk between 1 and 2 cell stages. At these stages the cytoplasmic flow into the cleaving embryonic cells carries the PAC DNA into the embryo. This process can be followed by the red colour of the DNA solution due to the phenol red.

After injection, around 25 embryos or less were placed in separate petri dishes containing new embryo medium. These were checked again for any deformities before they went in the incubator at 28°C. The same check was done every day, and the embryo medium was

replaced with fresh medium whenever necessary until the embryos achieved the 72 hpf required to score the *shd* phenotype.

3.2.3 Assembling the *shady* genomic region

3.2.3.1 Restriction analysis

After isolating the PAC DNA as described in the section 3.2.2.1, 600 ng of each PAC were digested with both Not I and Sal I restriction enzymes (Promega) together and independently. Not I is a 8 base-pair rare cutter and Sal I recognises 6 base-pair sequences. Each 20 µl digestion had 2 µl of enzyme (4 Units) and 2 µl of 10 x buffer D from Promega. The digestions were incubated at 37°C for two hours and then analysed by Pulsed Field Gel Electrophoresis (PFGE) using a CHEF-DRII (Clamped Homogeneous Electric Fields) (Biorad). Gels were made of 1% Fast lane agarose (Flowgen), run at 150 Volts for 27 hours in 0.5x TBE buffer with cooling to 15°C. A high molecular weight marker (Gibco BRL) was used in order to identify bands from 8.3 to 48.5 Kb. Around 300 ng of the marker was heated up to 65°C prior to loading the gels. Gels were post-stained with ethidium bromide until the bands were clearly seen in the printed picture (Sony video graphic printer). The same digests were also run on standard 0.8% agarose gel electrophoresis to detect bands smaller than 8 kb. These were run overnight at 20 volts for optimal resolution.

3.2.3.2. Non-radioactive hybridisation and detection of genomic DNA

Southern blots of the pulsed field gels produced were made according to Sambrook et al. (1989). In order to make a non-radioactive probe containing the z10985 marker a fragment of 304 bp containing the z10985 marker was amplified from genomic DNA and run on a 0.7%

agarose gel. The 304 bp band which had around 2μg of DNA was excised and diluted in MQ water. A ratio of 3 ml of water per gram of gel was used. Approximately 1μg of DNA was used in the labelling reaction. It was decided to employ the DIG High Prime DNA Labelling and Detection Starter Kit II (Roche) for the non-radioactive DNA labelling procedure. This Kit uses digoxigenin (DIG), a steroid hapten, to label DNA probes for hybridisation and subsequent chemiluminescence detection by enzyme immunoassay. The probes were generated with DIG-High Prime according to the random primed labelling technique. DIG High Prime is a specially developed reaction mixture containing digoxigenin-dUTP, and all reagents, including enzyme necessary for random primed labelling, premixed in an optimised 5 x concentrated reaction buffer. The use of a special form of DIG-11-dUTP enables easier and more efficient stripping of blots for rehybridisation with a second DIG-labelled probe. The hybridised probes are immunodetected with anti-digoxigenin-AP, Fab fragments and are then visualised with the chemiluminescence substrate CSPD, ready-to-use. Enzymatic dephosphorylation of CSPD by alkaline phosphatase leads to a light emission at a maximum wavelength of 477 nm which can be recorded with Lumi-Imager or on X-ray films. I have used X-ray film with exposure times in the range of 5 to 30 minutes and developed the film in a X-ograph Compact X2. The procedure for using this non-radioactive Kit was followed as described in the instruction manual (Roche cat.No.1585614).

3.2.3.3 End-cloning and sequencing of PAC inserts

This method is based on the fact that there are no Xba I sites in the PAC vector. We can therefore cleave the PAC clone with Xba I and clone the linearised vector attached to each end into suitable vectors; e.g. pUC19. We used the vector pGEM-4Z (Promega) (see Figure 3.1 and 3.2). To select for the inserts containing the PAC-ends both kanamycin and ampicillin

selections were used. The first selection was needed for the PAC-ends and the second for the plasmid vector, respectively.

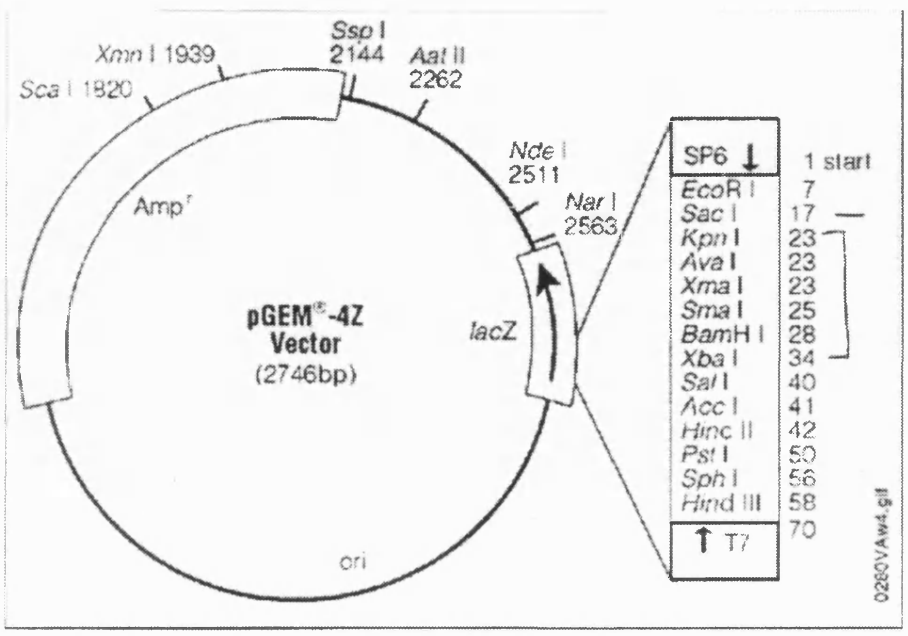


Figure 3.1. Diagram of the pGEM-4Z vector (Promega) showing its multiple cloning site (www.promega.co.uk)

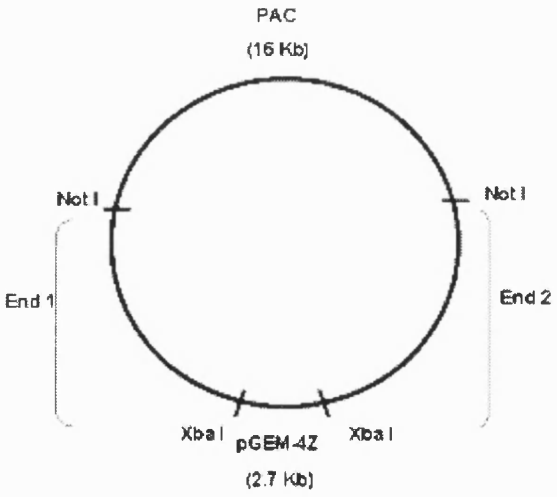


Figure 3.2. Diagram of the mixed vector containing both PAC-ends. The enzymes Not I and Xba I will release the PAC-ends for their size characterisation.

The procedure involved digesting 50-100 ng of PAC clone and 100 ng of pGEM-4Z with Xba I for 30 minutes at 37°C. The digestion was heat-inactivated at 65°C for 20 minutes and purified using Microcon-100 (Millipore) to remove salts. A standard ligation step took place in the same tube. Around 20 µl of the purified digestion was incubated with 2.5 µl of ligase buffer, 1µl of T4 ligase (Gibco/BRL) and 1.5 µl of MQ water for two hours at room temperature. Ligated vectors were electroporated into DH10B cells, expressed for one hour in 2 XYT medium and plated onto LB-agar plates with kanamycin (25 µg/ml) and ampicillin (50 µg/ml). This dual selection was necessary to select those clones that had both PAC ends and pGEM-4Z. The plates were left growing over night and on the next day appropriate clones were picked to grow minipreps for DNA sequencing. DNA isolation followed the protocol described earlier in section 3.2.2.1 using Plasmid-Safe ATP-dependent Dnase enzyme (Epicentre Technologies), with the exception that this time 2 ml cultures were used.

A 20µl restriction digestion including: 2µl Xba I (Promega), 2µl Not I (Promega), 2µl buffer D (Promega), 5µl DNA and 9µl MQ was made in order to check the sizes of each PAC-end. PAC-end sequencing could not be done using T7 and SP6 because these primers are present in both the PAC vector and the pGEM-4Z. Instead, M13-forward and M13-reverse primers were used since these are present in the pGEM-4Z vector but not in the PAC vector. Between 300-500ng of plasmid and 1µl of 10µM primer in a total of 6µl were sent for sequencing.

3.2.4 Identification of candidate loci for *shd*

The UK Human Genome Mapping Project - Resource Centre (UK HGMP-RC) makes available a tool named *NIX* (Nucleotide Identify X) which is intended to aid the identification of interesting regions in genomic or transcribed nucleic acid sequences. There are many useful

computer tools that can be used for this; however, none of them is completely accurate and it is useful to be able to compare the results from many programs that use different methods and have different strengths and weaknesses. *NIX* integrates many of these programs (including GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, Fgene, BLAST (against many databases), Polyah, RepeatMasker and tRNAscan) runs them on a sequence and displays their results side by side, thus identifying features detected by multiple programs. Rather than giving the user total control over how the programs are run by providing them with many poorly-understood choices, the NIX system selects reasonable defaults for all the programs, based on whether the sequence is genomic or transcribed, its species of origin and its size. This makes a vastly simplified interface for running a dozen or so programs on the specified sequence. Guidelines on how to use NIX efficiently were followed on line at http://www.hgmp.mrc.ac.uk/NIX/index.html.

3.3 Results

3.3.1 Screening P1 artificial chromosome genomic libraries

Pooled PAC libraries from the RZPD resource centre were screened by PCR using the marker z10985. According to the data obtained from linkage mapping (see Chapter 2) this marker showed a very tight linkage to the *shd* locus and the prediction was made that it should be only around 70 kb away from *shd*. This distance is within the range of one or two PAC clones and therefore we decided to screen this type of genomic library.

The primary screen results are presented in Figure 3.3. From the 33 pools screened, which represented 33 x 96 clones, 10 pools were positive for the presence of the z10985 marker (10, 11, 12, 13, 14, 19, 21, 22, 23, 33). From these, 8 pools were selected to carry on to the next screening step based on the fact that they reproducibly produced stronger, brighter and consistent bands. These selected primary pools were numbers 12,13,14, 19, 21, 22, 23 and 33.

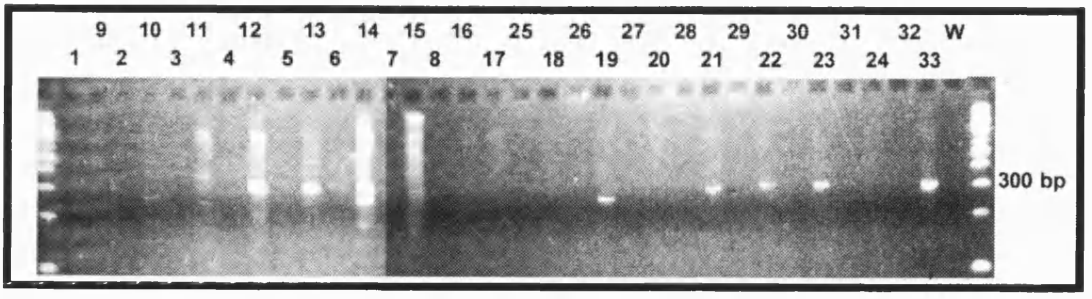


Figure 3.3. Primary PAC-pool screening procedure. 1% agarose gel showing the products of the primary screening with the marker z10985 (304bp). Primary pools tested are indicated above each lane. A water control was included in the last lane.

The secondary screening of these 8 pools of clones identified three plates that produced the consistent three bands necessary for the identification of the plate, row and column of the final clone. Figure 3.4 presents one example of such a gel.

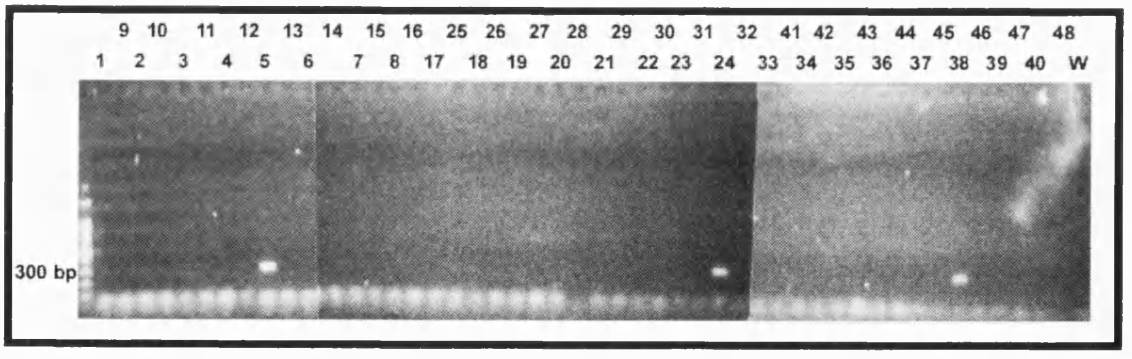


Figure 3.4. Secondary PAC-pool screening procedure. 1% agarose gel showing the products of the secondary screening on one of the plates that produced only 3 positive bands (5, 24 and 38) as required for identifying the final PAC clone. A water control was included in this experiment (W).

The identified clones were clone BUSMP706P14181Q2, which we will call PAC1 from now on; clone BUSMP706N10265Q2 which we renamed PAC2 and clone BUSMP706O16107Q2, which was renamed PAC3. All these three clones were from the same zebrafish PAC library 706, the vector is pCYPAC-6 and the host *E.coli* DH10B.

3.3.2 Rescuing the *shady* phenotype by genomic DNA microinjection

Injecting the three PACs into early developing mutant zebrafish embryos was used as a means to rescue the mutant phenotype and identify the critical region containing the *shd* locus (see Introduction). The results from these experiments are all summarised in the following three tables. The DNA concentrations inferior to 50 pg/nl were not considered for calculating the rescue rates. Only concentrations equal or superior to 50 pg/nl were efficient in producing rescue of the *shd* phenotype that could be scored at 72 hpf.

Table 1. PAC1 genomic DNA injection into 1-2 cell stage *shd* ^{ty82} zebrafish mutant embryos does not rescue iridophore phenotype

PAC1 injection experiment	*DNA concentration (pg/nl)*	*Number of injected embryos*	*Number of survivors at 72 hpf*	*Number of mutants*	*Number of rescued embryos*
1	68	58	58	13	0
2	68	146	109	30	0
3	50	88	70	25	0
4	50	105	76	23	0
TOTAL	>= 50	397	213	91	0

Table 2. PAC2 genomic DNA injection into 1-2 cell stage *shd* ^{ty82} zebrafish mutant embryos does not rescue iridophore phenotype

PAC2 injection experiment	*DNA concentration (pg/nl)*	*Number of injected embryos*	*Number of survivors at 72 hpf*	*Number of mutants*	*Number of rescued embryos*
1	50	500	205	38	0
2	50	150	78	26	0
3	50	150	122	26	0
4	55	160	148	30	0
TOTAL	>= 50	960	553	120	0

Table 3. PAC3 genomic DNA injection into 1-2 cell stage *shd* ^{ty82} zebrafish mutant embryos rescues iridophore phenotype.

PAC3 injection experiment	*DNA concentration (pg/nl)*	*Number of injected embryos*	*Number of survivors at 72 hpf*	*Number of mutants*	*Number of rescued embryos (%)*
1	68	13	13	2	0 (0)
2	25	129	113	25	0 (0)
3	50	153	104	34	6 (17.6)
4	50	256	102	37	9 (24.3)
5	60	300	271	69	4 (5.8)
TOTAL	>= 50	851	603	167	19 (11.4)

As an example of the extent of rescue obtained in experiments 3,4 and 5 from Table 3, Figure 3.5 shows some rescued embryos compared to their wild-type siblings. These rescued phenotypes are clearly outside of the range of iridophore phenotypes seen in *shd* ^{ty82} homozygotes. The strong *shd* allele *ty82* is characterised by having virtually no iridophores. The only exceptions I have encountered during the 4 years of this project were the presence of one tiny iridophore on the interior side of the eye, in a lateral patch, or in the dorsal stripe. Rescued phenotypes ranged from several individual iridophores to spots made of several individual iridophore cells or patches made of an even larger number of individual iridophore cells. Therefore it was concluded that these mutant fish were indeed partially rescued in respect to the iridophore lineage. Sometimes one eye or one lateral patch could look completely rescued. However, rescue was never to the extent that confusion with wild-type siblings was possible.

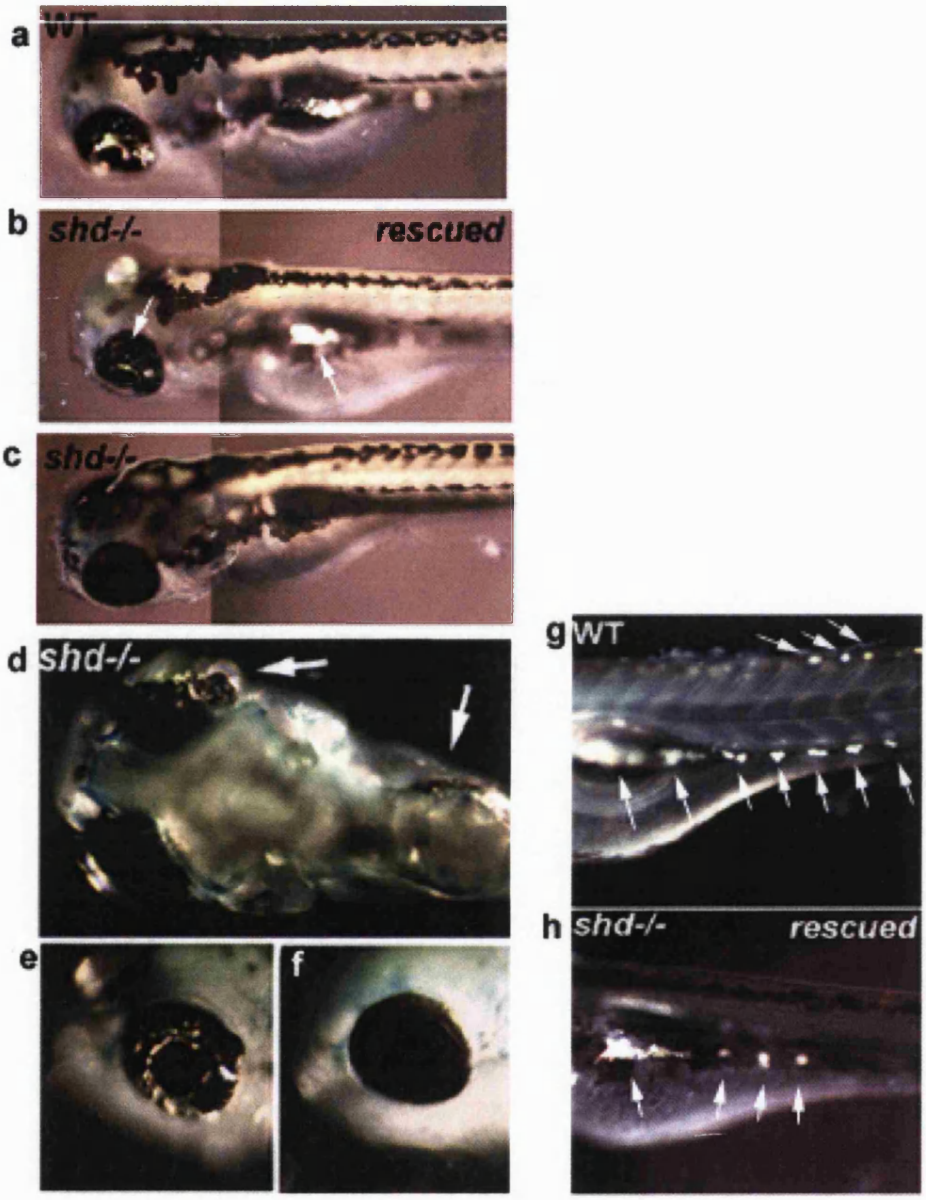


Figure 3.5. Phenotypic rescue of *shd* $^{ty82\ /\ ty82}$ mutants after PAC3 injections. (**a**) 72 hpf wild-type embryo; (**b**) same age *shd* $^{ty82/\ ty82}$ mutant embryo showing partial rescue of some iridophore spots in the eye and lateral patch after genomic injection with PAC3; *shd* $^{ty82/\ ty82}$ mutant embryo without rescued iridophores. These three embryos were siblings. (d) *shd* $^{ty82/\ ty82}$ mutant embryo rescued on the right side eye and lateral patch only (arrows). (e) detail of the rescued right side eye and (f) non rescued left side eye of the same embryo. (g) wild-type embryo showing part of dorsal stripe (top arrows), lateral patch region and part of ventral stripe with normal iridophore pattern (lower arrows). (h) sibling *shd* $^{ty82/\ ty82}$ mutant embryo showing partial rescue of lateral patch and three ventral stripe iridophores only.

3.3.3 Assembling the *shady* genomic region

In parallel to the PAC DNA injections referred to in this Chapter, we assessed how the three PACs overlapped, since they all shared the common marker z10985. In order to do that, restriction analysis using some rare cutter restriction enzymes was performed. The resulting fragments were run on pulsed field gel electrophoresis (PFGE) and to complement this analysis a probe containing the marker z10985 was synthesised and hybridised to the Southern blots of these PFGE products. The cloning and sequencing of some of the PAC ends also helped to position the three PACs correctly. At this stage the Sanger Centre started the zebrafish genomic project and was prioritising the sequencing of genomic contigs. It was therefore decided to send them the three PACs for shotgun sequencing. In this way, It was hoped that the bulk of information generated by the restriction analysis and PAC-end sequencing could be confirmed by the complete PAC sequences.

3.3.3.1 Restriction analysis

PFGE enabled restriction enzyme mapping of PACs 1-3 and demonstrated significant overlap (Figure 3.6).

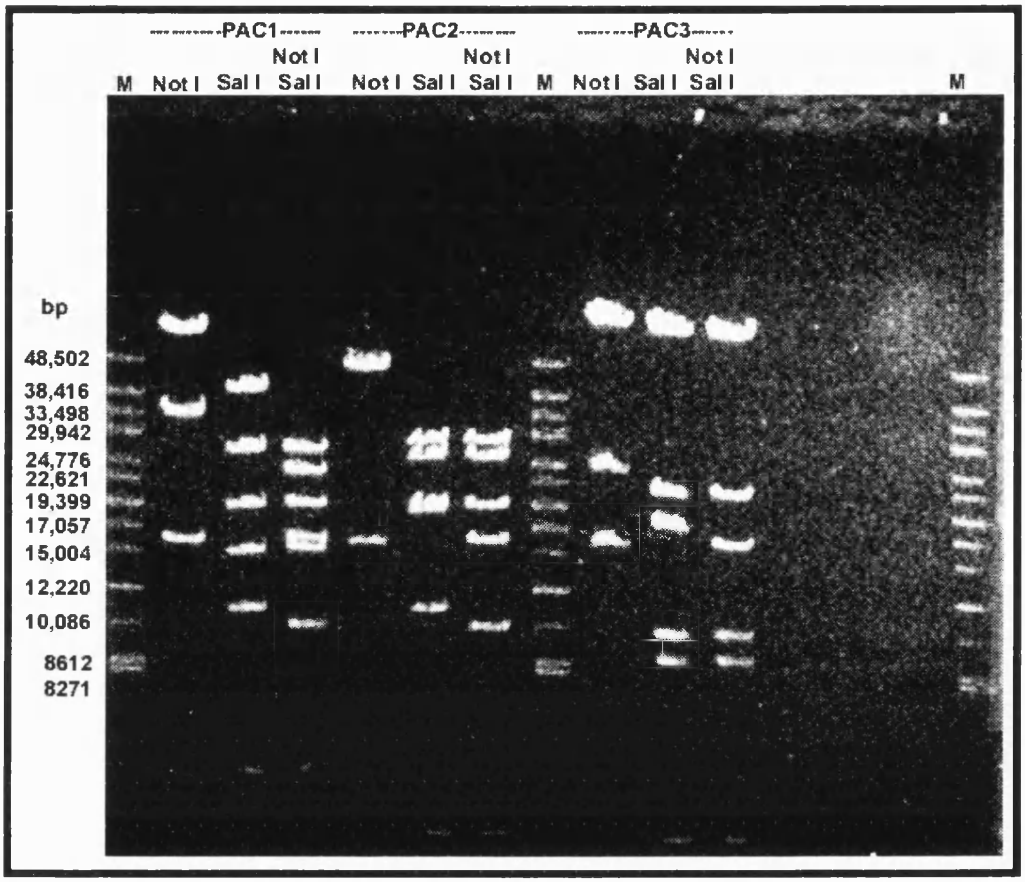


Figure 3.6. Pulsed Field Gel Electrophoresis of the Not I and Sal I digestion of PAC1, 2 and 3. It is clear that many bands have the same sizes for PACs 1 and 2. As an example, Sal I bands of around 25Kb and Not I/Sal I bands of the same size appear to be shared between PACs 1 and 2. All PACs share a 16Kb Not I band that corresponds to the PAC vector. (M) high molecular weight marker.

To detect some smaller fragments that might have run off the gel aliquots of the same digests were run on a 1% agarose gel. These results are presented in Figure 3.7 and show a number of fragments that were not visible in the PFGE results. In particular PACs 2 and 3 seem to have smaller fragments that if undetected would definitely hinder map assembly.

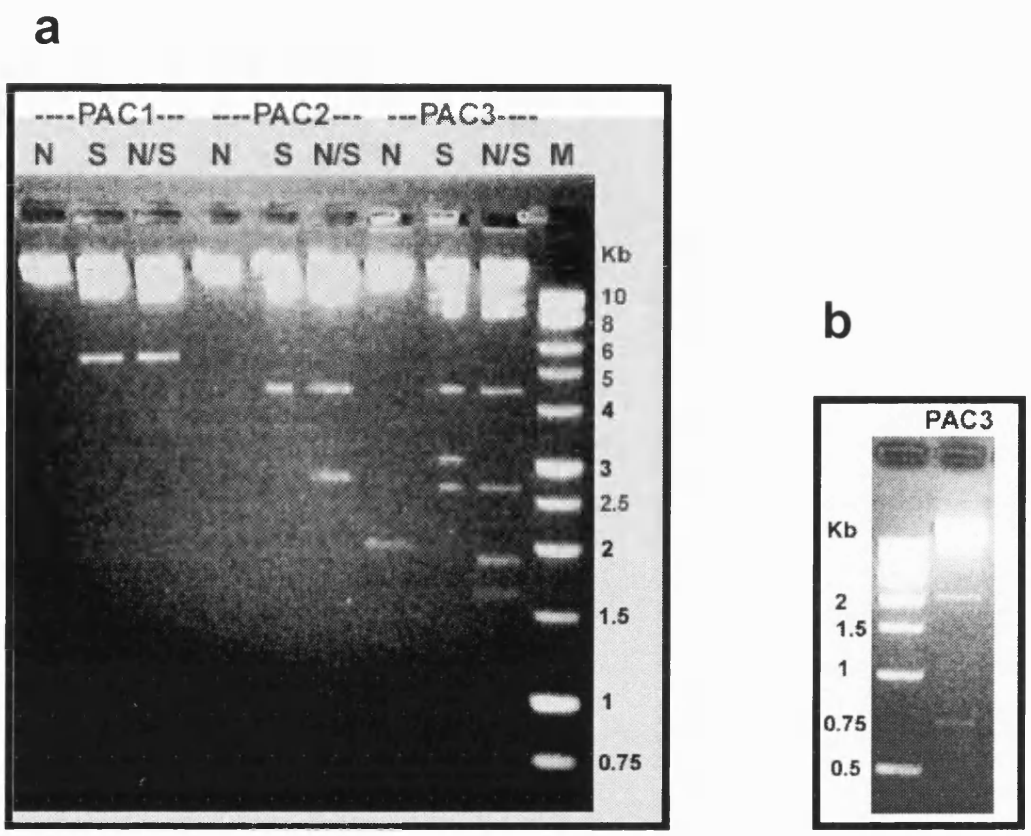


Figure 3.7. a) Not I and Sal I digestion of PAC1, 2 and 3 run on 1% agarose gel electrophoresis to detect fragments smaller than 10 Kb. (M) 1kb ladder. **b)** PAC3 digestion with Not I to visualise fragments smaller than 2Kb run on 1% agarose gel.

The interpretation of the size of each fragment showed in the Figures 3.6 and 3.7 is summarised in Table 4.

Table 4. Summary of the fragment sizes from the gel products of the restriction digests with Not I and Sal I presented in Figures 3.6 and 3.7 Similar size bands between PACs have the same colour

	Not I (kb)	Sal I (kb)	Not I / Sal I (kb)
PAC1	>50; 35; 16	38.5; 27.5; 19.3; 15; 11; 5.5	27.5; 24.7; 19.3; 16; 15; 10; 5.5
PAC2	>48.5; <48.5?;16	29.9; 27.5;19.3; 18?;11; 4.5	29.9; 27.5; 19.3; 16; 10; 4.5; 2.8
PAC3	>50; 24.7; 16; 2; 0.75	>50; 22; 17.5; 10; 8.6; 4.5; 3.2; 2.7	>50; 22; 16; 10; 8.6; 4.5; 2.7; 1.8; 1.6

The total size of PAC1 according to the sum of all the Sal I fragments is 118.3 Kb which is only slightly smaller than the size given by the addition of all the Not I/Sal I fragments 119.5 Kb. The same exercise for PAC2 indicates a much larger discrepancy between the total size calculated with the Sal I fragments (92.2 Kb) and the size given by the double digest (110 kb). In fact this discrepancy might be telling us that there is another Sal I fragment of around 18 Kb that has not separated from the 19.3 Kb fragment well enough for us to detect it. Consistent with this, the Sal I 19.3 Kb band is wider than the others suggesting that this might be the case. The same discrepancy happens with the sum of Not I fragments from PAC2. The wide >48.5 Kb fragment must be a double band formed by two fragments of slightly different sizes that did not separate well enough after electrophoresis. In relation to PAC3 the difference between the addition of Sal I fragments (118.5 Kb) and the double digest fragments (117.2) is very small, indicating that our estimate sizes are probably nearly correct. These estimates were assuming the larger fragments are 50 Kb however, these can be larger. Attempts to use a molecular marker that allowed estimation of fragments larger than 48.5 Kb were not

successful. To enable construction of a restriction enzyme map, this digestion data was supplemented with Southern blot analysis and non-radioactive hybridisation presented in the next section.

3.3.3.2 Non-radioactive hybridisation and detection of genomic DNA

Hybridising a probe containing the marker z10985 (after excluding a region with a stretch of CA repeats) to a membrane obtained from the gel present on Figure 3.6 should tell us which bands contained this marker. After identifying the overlapping bands anchored by z10985 we expected to be able to make better predictions on which bands are also likely to contain the *shd* gene. These results are presented in Figure 3.8

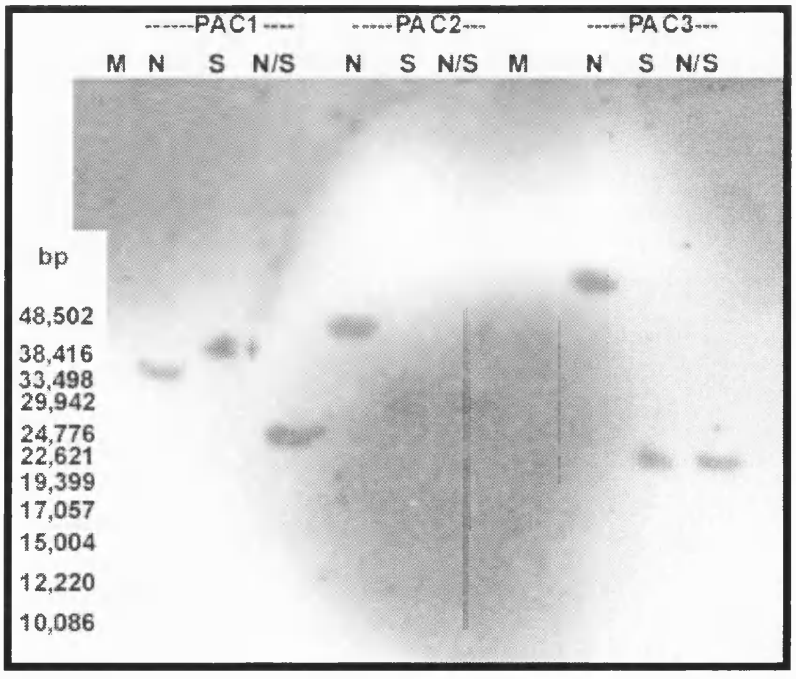


Figure 3.8. X-ray film with hybridisation of the probe containing the marker z10985 without CA repeats to the Southern blot of the gel presented in Figure 3.6.

Overlaying the x-ray film (Figure 3.8) to the original gel (Figure 3.6) allowed identification of the bands containing the marker z10985 (see Table 5).

Table 5. Identification of the Not I and Sal I restriction fragments that contain the marker z10985.

	Not I	Sal I	Not I / Sal I
PAC1	35 Kb	38.5 Kb	24.7 Kb
PAC2	>48.5 Kb	29.9 Kb	29.9 Kb
PAC3	>50 Kb	22 Kb	22 Kb

Using all the mapping data available, plus knowledge of the cloning sites in the PAC vector (Not I present at both ends of insert and Sal I only at SP6 end) it was possible to suggest a map of the region covered by the three PACs (Figure 3.9). It is perhaps worth noticing that the vector backbone is 16 Kb long. This fragment is isolated when cutting with Not I but it is only cut at one end when Sal I is being used on its own. So, for instance in PAC 1 the last Sal I fragment on the right end side has around 40.7 Kb (24.7 plus 16) which I estimated as 38.5 Kb in Table 4. In PAC 2 there is a sal I fragment of 18.8 Kb produced by the vector backbone plus the last 2.8 Kb mixed fragment. For PAC 3 it is possible to assume that the 18 kb Sal I fragment (estimated as 17.5 Kb in Table 4) is probably the result of the vector backbone plus the 1.8 Kb end fragment (see Figure 3.9).

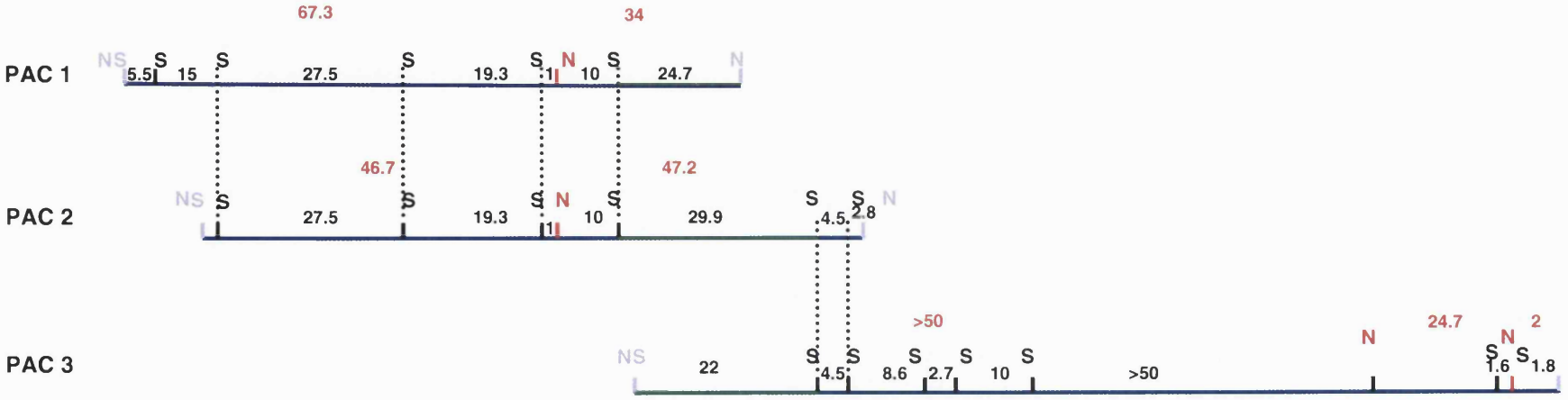


Figure 3.9. Map of the genomic region containing the three PAC inserts identified with the marker z10985. The fragments that hybridised to the probe z10985 are presented in green. Restriction sites and fragment sizes are represented in red for Not I and in black for Sal I. PAC vector restriction sites are represented in purple.

This map is the only one consistent with all the data. The only ambiguity concerns the precise order of the Sal I fragments to the right of the 4.5 Kb Sal I fragment in PAC3. The sizes for PACs 1 and 2 are unambiguous. The order of PAC3 fragments right of the zone of overlap with PACs 1 and 2 was indeterminate, but became clear after sequencing of the contig.

3.3.3.3 End-cloning and sequencing of PAC inserts

The sizes of the cloned PAC-ends were checked with both Xba I and Not I restriction enzymes. It was expected to see four bands in each digest : two correspond to the PAC and PGEM-4Z vector backbones (16 kb and 2.7 Kb, respectively) and the other two bands for each of the PAC ends. However, the mapping data obtained in the previous section showed multiple Not I sites near one of the ends of PAC3. It was therefore possible that more than two bands might be found for PAC3 end-clones resulting from Not I sites contained in the PAC-end clone.

Gel electrophoresis of the Not I/Xba I restriction digests identified the size of the end-clones for each PAC (Figure 3.10). As expected PACs 1 and 2 had 4 fragments each and PAC3 had 6, from which 2 were Not I fragments. Table 6 summarises these fragment sizes and accounts for the Not I fragments present in PAC3.

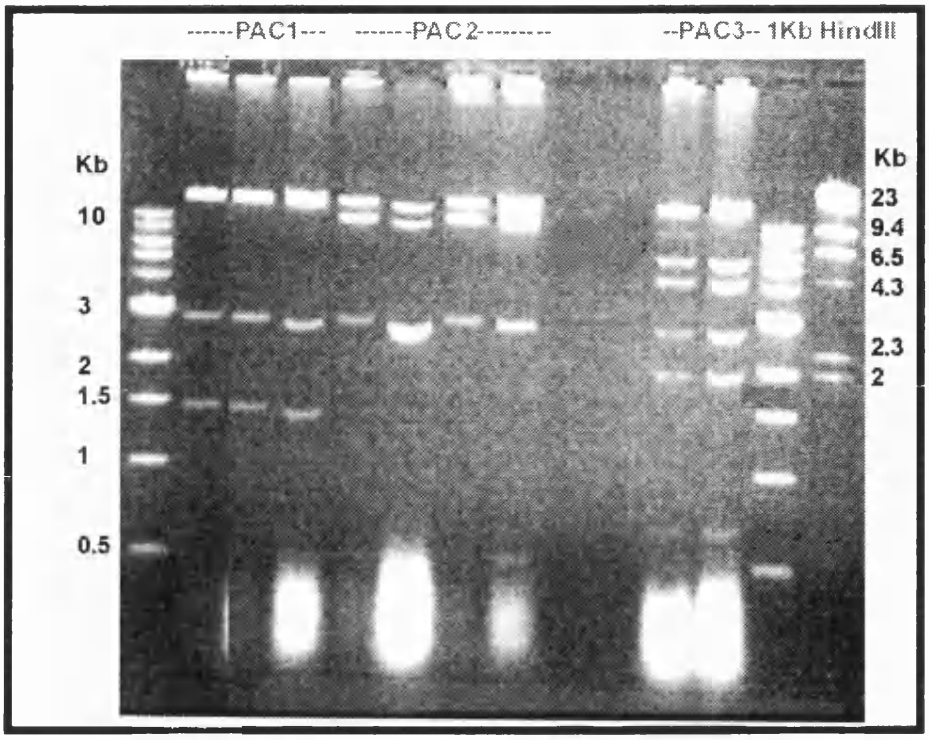


Figure 3.10. PAC-end clones. 3 minipreps for PAC1, 4 to PAC2 and 2 for PAC3 were doubly digested with Not I and Xba I to identify the size of the PAC-end clones. PAC vector is 16Kb and PGem-4Z is 2.7Kb. 1KB ladders are flanking the gel and a Hind III ladder is shown on the right side of the gel.

Table 6. PAC end-clones. Fragment sizes produced by the Not I/ Xba I digestions shown on the Figure 3.10 (above). Not I fragments for PAC3 are shown in the gel from Figure 3.7b.

	Xba I / Not I fragments (Kb)	Not I / Not I fragments (Kb)
PAC1	1.5; <0.5	-------
PAC2	11; 0.5	-------
PAC3	5.5; 4.3	2; 0.75

At this stage we had expected to identify unambiguously the right and left ends of each PAC on the restriction enzyme map presented in Figure 3.9 by making probes with each of the PAC end-clone fragments. This would be an independent confirmation of the genomic map that was presented before. However, in practice we chose an alternative strategy : the opportunity to have the contig sequenced by the zebrafish genome sequencing project was an attractive option. Thus, the PAC end-clones were sequenced for comparison with the Sanger Centre sequences. The end clones were sequenced with M13 primers. However, only the M13 Forward primer worked, since there is an unreported M13 reverse primer sequence in the middle of the PAC vector backbone. From directly sequencing PAC2 with T7 primer some sequence was obtained from the other end of PAC2. Comparisons between these PAC-end sequences with those released by the Sanger Centre are presented in the next section where a new map of the *shd* genomic region integrates all the sequencing data.

3.3.4. Identification of candidate loci for *shd*

As soon as the draft sequences for the three PACs became available from the Sanger Centre they were analysed with the NIX package for identification of potential coding sequences. The first PAC sequence to be completed was PAC3. This was fortunate because PAC3 injection rescued *shd* mutants and hence contained the *shd* locus. The Sanger Centre gave the following accession numbers to the three PAC sequences: PAC1 AL592543; PAC2 AL672149 and PAC3 AL596022. These are available in the genome database and have sizes of 99495Kb; 15645Kb and 121941 Kb, respectively.

Clicking on the boxes/genes given in the NIX output (Figures 3.11-3.13) reveals details of the features identified. Here we summarise the most reliable features identified on each PAC. The output for PAC1 (Figure 3.11) indicated that it contains a novel gene similar to

IVD (HUMAN P26440 Isovaleryl-CoA dehydrogenase, mitochondrial precursor (EC 1.3.99.10) and a fragment of a novel gene similar to ITPKA (inositol 1,4,5-trisphosphate 3-kinase A) also known as IP3K. According to the Swissprot database the 5' end of IP3K was situated at the end of the sequence counting from the upper strand (above the sequence line in thick green). The IVD predicted gene was localised around 63856-77362 bp but it was on the opposite strand (below the sequence line). The EMBL database indicated that PAC1 and PAC3 overlapped from 85557 bp until the end of the PAC1 clone. This overlap included the IP3K predicted 5' fragment which spread from around 90230-93334 bp according to the BLAST programs. The promoter region of this gene was estimated to start at position 78778 bp and the first exon at 80030 to 80476 (according to Genscan software).

The results released for PAC2 sequence (Figure 3.12) indicated the presence of a potential gene in the upper strand (above sequence line). The BLAST results showed this gene to be similar to human 1D- Myo-Inositol Trisphosphate 3-Kinase A (EC 2.7.1.127), also known as Inositol 1,4,5-Trisphosphate 3-Kinase, IP3K or IP3-kinase, the same gene as the identified fragment in PAC1. NIX identified four exons from this gene over which most programs agreed. Three exons were internal and one was predicted to be terminal. The first one spanned the sequence from 3924-4128 bp and had a probability of 0.968 of being correct. The second one was between 5889-5987 bp and had a probability of 0.665. The third one was between 8088-8159 bp with a 0.506 probability and the terminal one spread from 11846-12052 bp and had a probability of 0.651.

The PAC3 sequence (Figure 3.13) was predicted to contain two potential genes, one over a shorter region on the lower strand (below the sequence line) and another over a very large region on the upper strand (above the sequence line). The gene predicted on the lower strand was very similar to human 1D- Myo-Inositol Trisphosphate 3-Kinase A (EC 2.7.1.127) and was also on the overlapping regions of the other two PACs. Once again,

because several programs detected this gene it is considered a good prediction. The larger gene predicted on the upper strand of the sequence was similar to two human genes, Anaplastic Lymphoma Kinase (ALK) and Leukocyte Tyrosine Kinase (LTK).

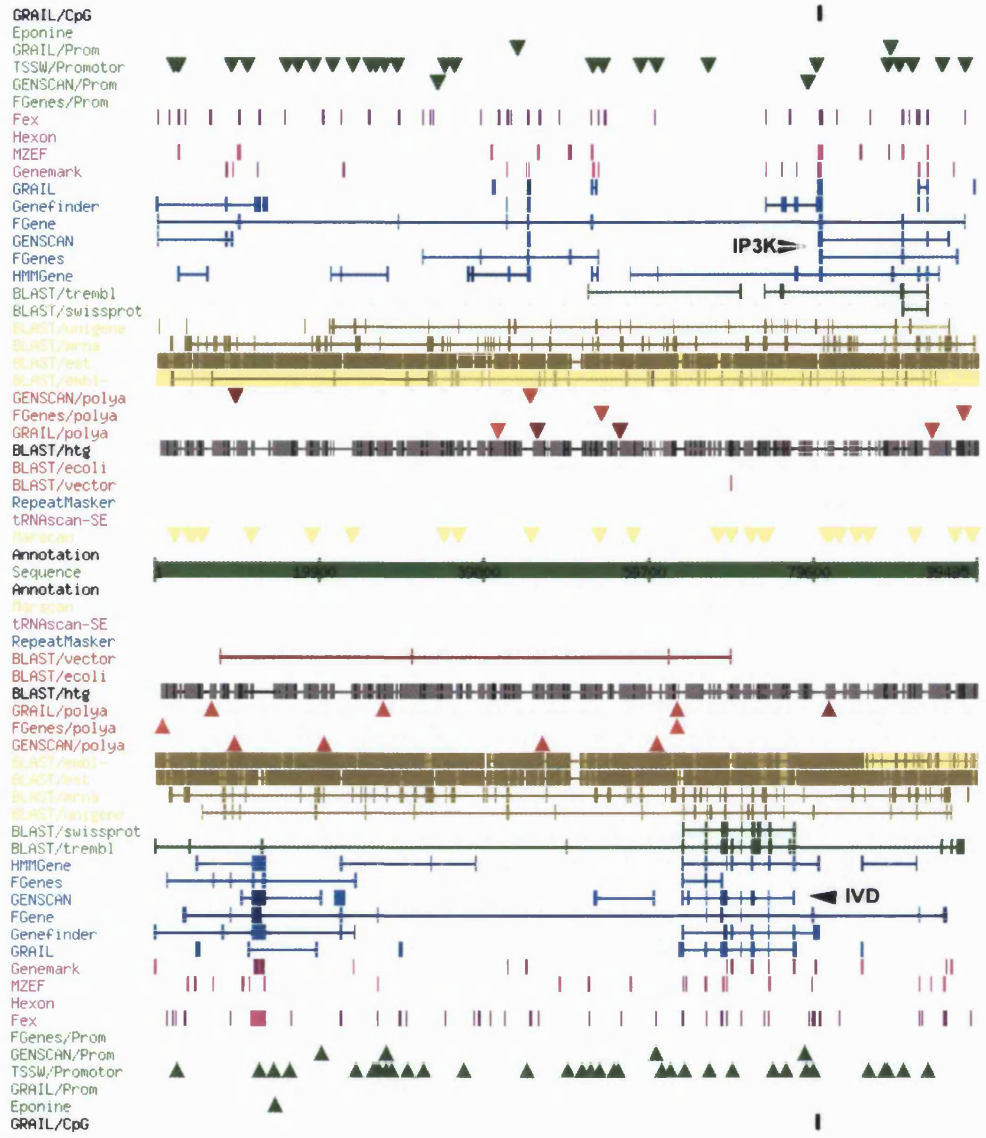


Figure 3.11. NIX output for the analysis of PAC1 sequence. The programs run in this package are listed on the left side. They are grouped by colour according to similar purposes. The blue colour codes for programs that look for potential genes. The different shades of blue indicate the quality of the prediction. The more intense the blue the better the prediction. The rectangular boxes indicate potential exons linked by lines if belonging to the same predicted gene. If different programs agree in predicting one given exon it makes it a very good prediction. Clicking on the boxes on-line will give the raw output produced by each program. Clicking on the lines will give information on the start and end positions of that predicted gene. The green colour is given for the BLAST programs Swissprot and TREMBL. It indicates that the coding regions predicted by the programs in blue are actually similar to sequences already present in the databases. The genes identified in this way are indicated in black. It is possible to access the raw BLAST results or a graphical view of the BLAST matches. Other BLAST programs in yellow, red and black include different EMBL databases such as mRNA, EST, HTG (High Throughput Genome Sequences). There are also BLASTs for the STS EMBL database, E. coli database, and a vector database. The sequence orientation is the same as used in the restriction enzyme map from Figure 3.9.

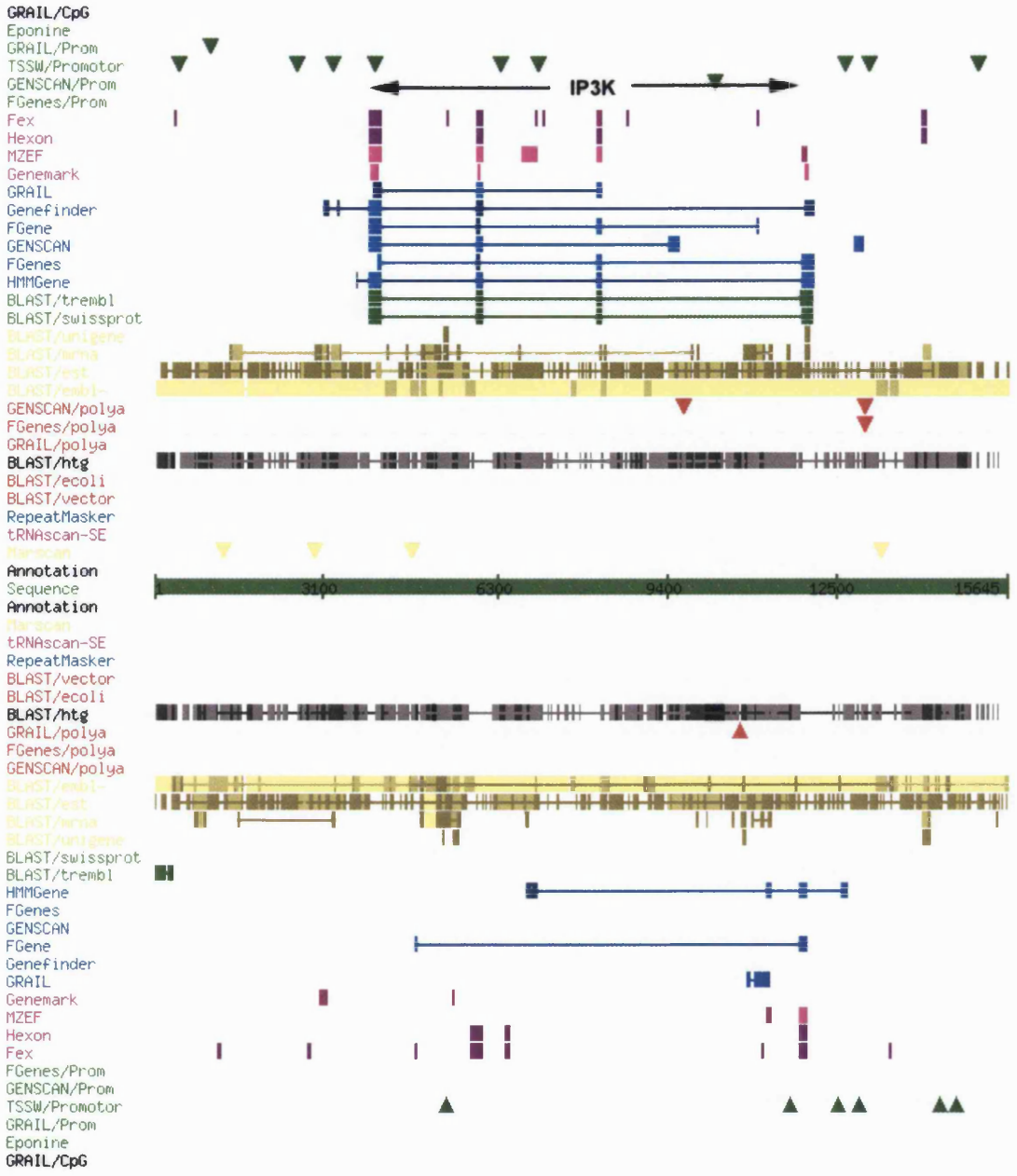


Figure 3.12. NIX output for the analysis of PAC2 sequence. Programs on the left are grouped by colour depending on their purpose. Rectangular boxes/sticks that are vertically aligned mean that different programs agree over that particular feature. See legend from Figure 3.11 for a detailed description of the colour code. The sequence orientation is the same as used in the restriction enzyme map from Figure 3.9.

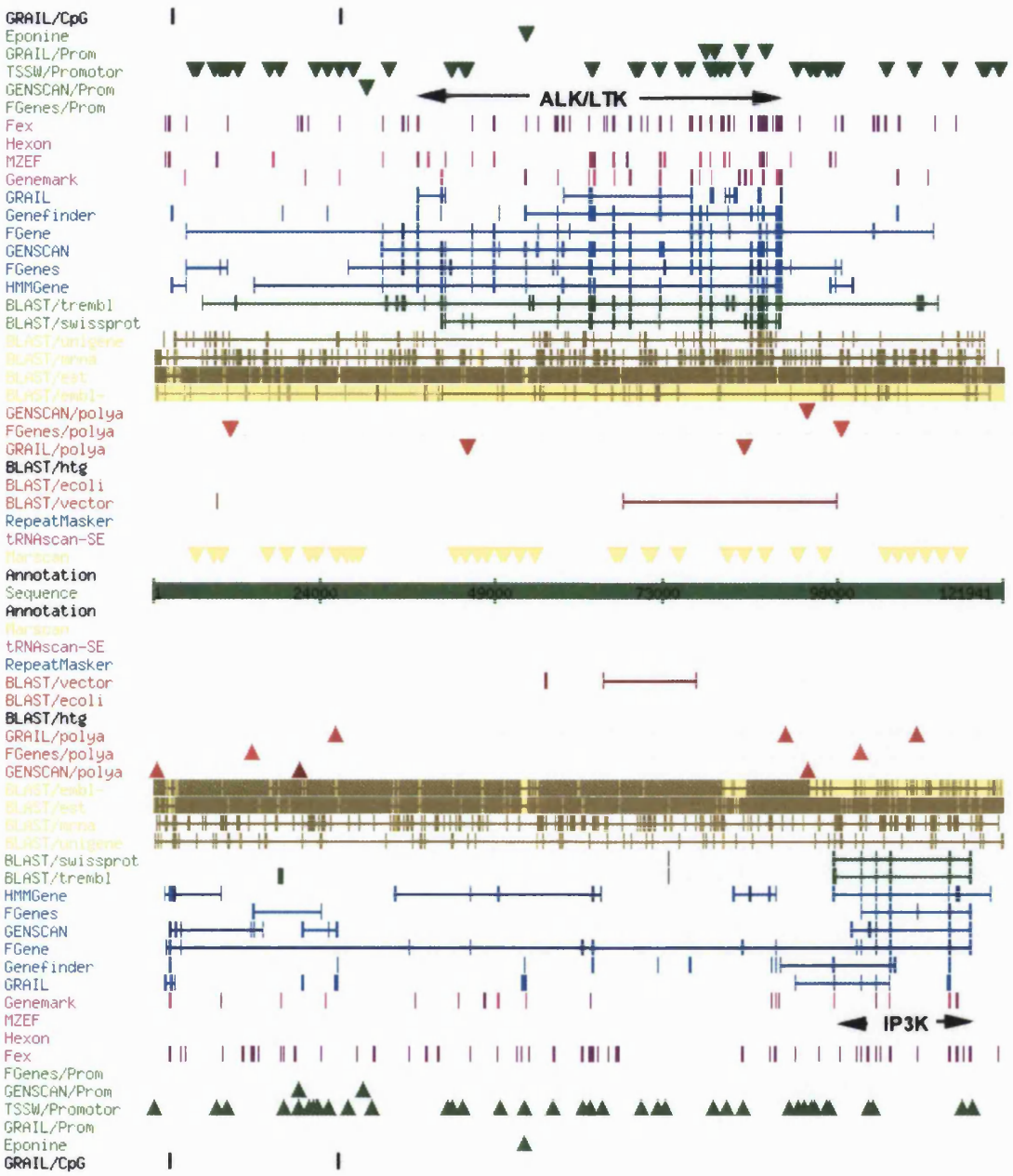


Figure 3.13. NIX output for the analysis of PAC3 sequence. Programs on the left are grouped by colour depending on their purpose. Rectangular boxes/sticks that are vertically aligned mean that different programs agree over that particular feature. See legend from Figure 21 for a detailed description of the colour code. The sequence orientation is the opposite to the one used in the restriction enzyme map from Figure 3.9 and to the one used in Figures 3.11 and 3.12.

BLASTs between two sequences (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html) were performed to confirm the overlapping regions between PACs, localise the PAC-end sequences (see section 3.3.3.3), and most importantly to confirm where the marker z10985 (accession number: G39855) mapped.

Based on all these results it was possible to construct a map showing how the three PACs overlapped based on the sequencing results (Figure 3.14). The first feature that stood out from this new analysis was the size of PAC2. The Sanger Centre PAC2 sequence was much shorter than the insert size deduced by restriction enzyme analysis in section 3.3.3.2 in this Chapter. The estimated size of PAC2 based on restriction fragment analysis with Not I and Sal I was around 94000 bp compared to 15645 bp of released sequence. Apart from this aspect both maps are highly consistent, both in terms of the sequences obtained and the observed/predicted restriction fragments.

As expected, the PAC end-sequences available mapped to the ends of the PACs next to a Xba I site (see section 3.2.3. for explanation). Their positions further confirmed the overlap between the PACs (see Figure 3.14). It is important to notice that the left end-sequence of PAC2 mapped to the positions 31345-31971 bp of PAC1 supporting the idea that the sequence obtained for PAC2 by the Sanger Centre was incomplete.

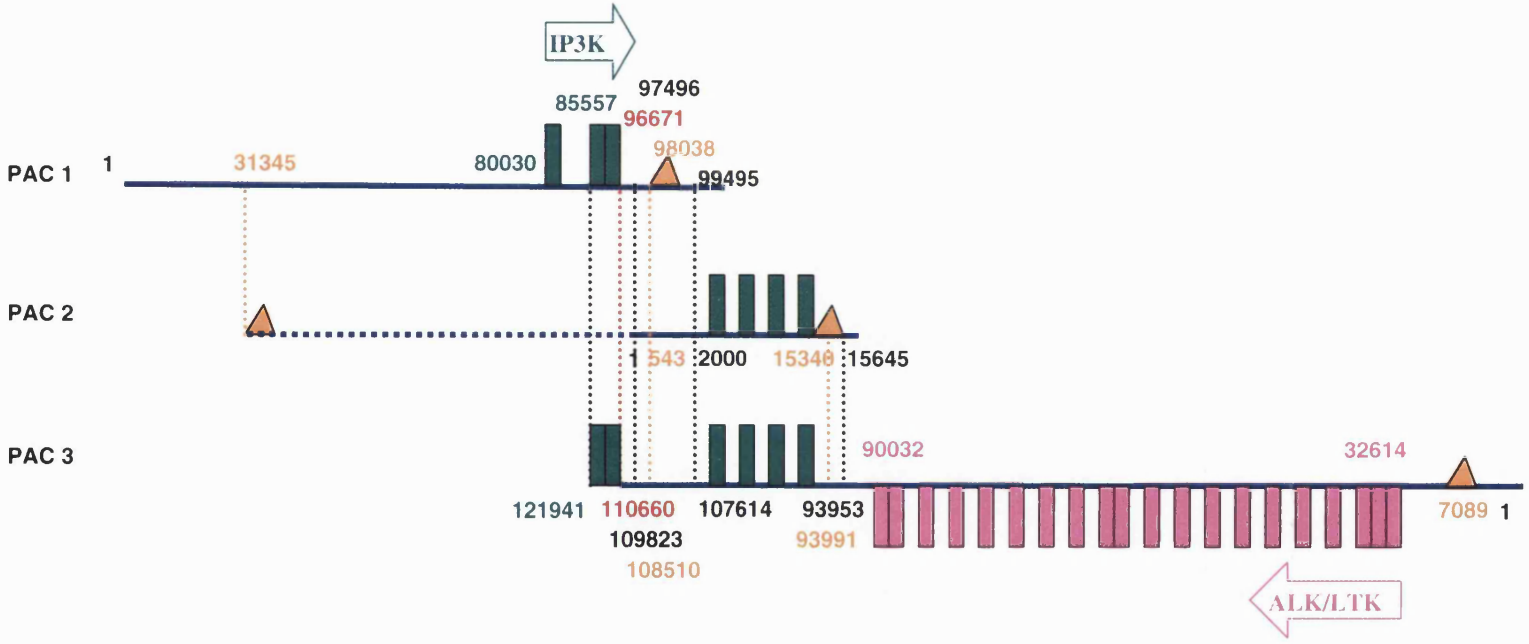


Figure 3.14. Map of the *shd* genomic region containing the three PACs based on NIX analysis and end-sequencing results. The red colour marks the position of the marker z10985. Orange triangles sign the positions of the end-sequences and their correspondent positions on the other PACs are also in orange numbers. The blue dotted line represents the region of PAC2 that was absent from the Sanger Centre data. The rectangular shapes represent exons of predicted genes, IP3K (green) and ALK/LTK (pink). The arrows show the orientation and beginning of the predicted transcription site. Numbers indicate genomic positions in base pairs (bp).

Since only PAC3 injection could rescue the *shd* $^{-/-}$ phenotype (see section 3.3.2) we focused on the two genes predicted to be in this clone. Whether PAC3 contained the full IP3K gene is ambiguous because the first exon was not detected as such by any NIX program except the HMMgene package. This ambiguity is exemplified by considering the prediction of the Fgenes and HMMgene softwares. The IP3K predicted gene contained either six or seven exons spread from 101479-117276 bp (Fgenes) or from 97541-120209 bp (HMMgene). The probability given for each of the exons and their boundaries also varied between programs. Similarly, the BLAST results were not very robust either; the prediction was only marginal and referred to a fragment that ranged from 97547-117276 bp.

In contrast, the presence of a full-length ALK/LTK gene on PAC3 was more robustly predicted although the quality of these predictions again varied for each predicted exon. Nevertheless, most programs seemed to agree that the gene has around 20 exons covering a genomic region starting around 32614 bp and ending at 89061-90032 bp. There were identical BLAST results from Swissprot and EMBL databases recognising a good similarity to the ALK and LTK human genes. The genomic region identified by these databases spread from 41203-89804 bp meaning that the predicted gene probably has a specific novel 5'end that is not recognised by the genome database.

3.4 Discussion

We chose to use PAC injections of mutant embryos as a screen to narrow down the genomic region of interest. At the time there were no published examples of this approach, so one concern was whether we could obtain broad distribution of such a large construct in the embryo. There were at least two factors to take into account; the first being the mosaic distribution of the injected genomic DNA in relation to the size of the target population of cells. The second factor was the concentration of the genomic DNA with respect to the number of cell divisions that happen from early blastomere until the iridoblast formation.

After a few trials the volume injected in this work was around 10 nl per embryo at a concentration of 50pg/nl which results in 500 pg of genomic DNA per embryo. This concentration seemed to be the minimum required to produce rescue of iridophore cells at 72 hpf. *floating head* (*flh*) mutants (Yan et al., 1998) were also rescued with genomic DNA injections but the authors do not mention the volume injected into each embryo, nevertheless the concentration that worked better was also 50pg/nl and the percentages of rescued embryos are comparable. The fact that the *flh* phenotype is visible at 22 hpf compared to the much later time (72 hpf) for scoring *shd* mutants may account for their slightly better results. In another study where genomic microinjections were used to rescue mutant phenotypes the concentrations used ranged from 40-60 pg/nl (no mention of volume injected again) and the phenotype was scored at 48 hpf (Guo et al., 2000). This last study did not provide percentages of rescue either. Finally, there was one other work (Lyons et al., 2002) where embryos were injected with only 100 pg of BAC DNA. This amount was enough to rescue the bloodless phenotype of *vlad tepes* at 21 hpf (Lyons et al., 2002). Perhaps this can be explained by the number of blood cells at 21 hpf being much higher than the number of iridophore cells at 72 hpf.

The two genomic maps presented in this Chapter were generated by applying two completely independent approaches, restriction fragment analysis and sequencing, and gave very similar results. This result strongly validates the genomic analysis of this region. The sizes of the PAC clones were slightly overestimated by the restriction analysis approach, but the only major disagreement in size refers to PAC2. It is believed that the reason for the huge discrepancy in size of the PAC2 insert was probably a deletion that occurred in the clone used for sequencing. The facts that (1) the marker z10985 is not present in the PAC2 sequence and (2) the left PAC2 end-sequence overlaps with a region around position 31 Kb of PAC1 (see Figure 3.14) strongly support this argument. Taking this left end as the start of PAC2 the correct insert size would be close to 80 Kb or 96 Kb including the vector. This size is within the predicted size range of 92 -110 Kb proposed from the restriction analysis alone.

Combining the results from the genomic rescue with PAC3 with the data from the genomic maps it is possible to identify a high probability region within which the *shd* locus is likely to lie. PAC2 did not rescue the *shd* mutant phenotype therefore arguing strongly against IP3K as a candidate. However, conceivably, PAC3 might have a functional IP3K gene, whilst that in PACs1 and 2 is truncated. This, however, seems unlikely because NIX packages were unable to find a high probability initial exon or promoter region for IP3K in PAC3. Such a region was only present in PAC1 starting around position 78778 bp.

In contrast, the other strongly predicted gene within the critical region was a much better candidate. This gene encodes a receptor tyrosine kinase (RTK) similar to both mammalian ALK and LTK. Further support for the proposition that this RTK might be *shd* arises from consideration of : a) gene size; b) synteny; c) gene function. In Chapter 2 we concluded that the marker z10985 and the *shd* locus lay about 74Kb apart. Whilst the marker lies within IP3K, the 5'end of the ALK gene lies around 80 Kb away (see map on Figure 3.14).

In the Discussion of Chapter 2 it was argued that there was good synteny between the region where *shd* maps on LG 17 and a fragment of human chromosome 2p13 - 2p25. It happens that human ALK lies within this region (position 2p23) in support of its candidacy. LTK maps to human chromosome 15q15.1-15q21.1. So far there are only two zebrafish genes on LG 17 that have potential orthologues in human Chromosome 15. However, one of these is IVD, a gene predicted to be in PAC1 and therefore very close to the *shd* critical region. Consequently, syntenic arguments support the identity of the PAC3 RTK as an orthologue of either human ALK or human LTK. These arguments, however, also apply to IP3K as this gene maps to the same human chromosome as LTK (15q15.1-15q21.1). We, therefore believe that this region of LG 17 has genes that are syntenic with human chrmosomes 2 and 15, which does not help us find the right candidate for the *shd* locus.

Both ALK and LTK belong to the insulin receptor subfamily of receptor tyrosine kinases and have 57% amino acid identity and 71% similarity (Morris et al., 1997). This family includes other RTKs with well known developmental functions, such as the Insulin growth factor receptor I (Igf1r), mainly involved in embryonic and post-embryonic growth e.g.(Maitre et al., 1995; Pera et al., 2001) but also with an anti-apoptotic effect in tumorigenesis (Kuroda et al., 1998).

ALK was first identified as a fusion product with the *Nucleophosmin* gene in the t(2;5)(p23;q35) translocation in anaplastic large cell lymphoma (ALCL) (Morris et al., 1994). Therefore the majority of the work done on ALK explores the role this gene plays in cancer (e.g. Agarwal et al., 2002; Au and Liang, 2002; Bai et al., 2000; Bridge et al., 2001; Cook et al., 2001; Duyster et al., 2001; Evens and Gartenhaus, 2003; Gascoyne et al., 1999; Greenland et al., 2001; Hernandez et al., 2002; Pulford et al., 2001; Zamo et al., 2002). Not much is known about the normal function of ALK except that its murine transcripts are confined to the nervous system and include several thalamic and hypothalamic nuclei; the trigeminal, facial, and acoustic cranial ganglia; the anterior horns

of the spinal cord in the region of the developing motor neurons; the sympathetic chain; and the ganglion cells of the gut (Morris et al., 1997). It is noteworthy that many of these tissues actually belong to the peripheral nervous system and contain cells derived from the embryonic neural crest. However, there are no reports of ALK expression in melanocytes, the only type of pigment cell present in mammals.

The LTK gene, also known as protein tyrosine kinase-1 or TYK1, was first isolated from a mouse pre-B lymphocyte cDNA library and Northern analysis revealed expression of the gene in thymus, spleen and kidney (Ben-Neriah and Bauskin, 1988). The human LTK gene product was identified in placenta and hematopoetic cell lines (Kozutsumi et al., 1993). Toyoshima et al (Toyoshima et al., 1993) cloned a set of cDNAs representing differently spliced human LTK mRNAs. These cDNAs predicted a truncated receptor protein that lacked the tyrosine kinase domain and a soluble receptor protein that had neither a transmembrane nor a tyrosine kinase domain. A cDNA clone containing the complete open reading frame demonstrated that the extracellular domain of the receptor was larger than previously predicted by Ben-Neriah (1988). Thus, the LTK gene produces not only the putative receptor tyrosine kinase for an unknown ligand but also multiple protein products that may have different functions. Recently, however, it has been suggested that LTK may function in a survival signalling pathway (Honda et al., 1999; Ueno et al., 1995; Ueno et al., 1996).

Although it is possible to state, based on the results from this Chapter, that the RTK similar to both ALK and LTK is the most likely candidate for the *shd* locus present in PAC3. It will be crucial to test this prediction by doing laboratory work on these two candidate genes. The fact that mammalian ALK expression has been reported in neural crest derivatives is encouraging.

Chapter 4

Molecular characterisation of the *shady* locus

4.1 Introduction

In the previous Chapter an ALK/LTK-related Receptor Tyrosine Kinase (RTK) was introduced as a candidate for the *shd* gene product. These structurally similar RTKs belong to the insulin receptor superfamily of RTKs (also known as type II RTKs). Together, based on their primary structure, they form a novel subfamily of the Insulin receptor family of RTKs (Iwahara et al., 1997; Robinson et al., 2000). A wide range of processes involving intercellular communication are mediated by RTKs and their signalling pathways. The growing list of processes regulated by these receptors across the phylogenetic tree is very broad including the induction of cell fates, guidance of cell and axon migration, cell proliferation, differentiation and survival (Loren et al., 2001).

RTKs are composed of three domains: an extracellular ligand-binding domain, a single membrane-spanning domain and a cytoplasmic catalytic domain displaying the highest level of conservation (Yarden and Ullrich, 1988). Ligand binding to the extracellular domain induces activation of the kinase on the cytoplasmic side, which initiates the intracellular signalling. The activated RTKs phosphorylate themselves and the cytoplasmic substrates, leading to activation of a number of downstream signalling molecules, and ultimately inducing changes in gene expression and the phenotypic state of the cell (Fantl et al., 1993; Van der Geer et al.,1994). Phosphorylation of distinct tyrosine residues of the activated receptor creates binding sites for Src homology 2 (SH2) and phosphotyrosine binding (PTB) domain-containing proteins. Molecules recruited via these binding motifs are either enzymes that are activated by subsequent tyrosine phosphorylation, such as Src and phospholipase Cγ, or adaptor molecules that link the RTK activation to downstream signalling pathways. One important adaptor protein complex is the Shc-Grb2-Sos complex that couples RTKs via Ras to the extracellular regulated kinase (ERK) /mitogenic-activated

protein (MAP) kinase pathway which is crucial for RTK-induced cell proliferation (Zwick et al., 2001).

RTKs are not only key regulators in normal cellular processes but are also critically involved in the development and progression of human cancers and other hyper-proliferative diseases when they become constitutively activated. This can happen by several processes including deletions or mutations within the extracellular region or alterations of the catalytic domain, especially the ATP-binding motif. Mutations in the transmembrane domain have also been shown to lead to ligand-independent kinase activation (Zwick et al., 2001).

ALK (Anaplastic Lymphoma Kinase) owes its name to the type of human cancer where it was first identified (Morris et al., 1994). A translocation fused the *ALK* catalytic domain to the *Nucleophosmin* (*NPM*) gene which resulted in a chimeric constitutively activated tyrosine kinase that is expressed in a deregulated and ectopic manner, both in terms of cell type (lymphoid) and cellular compartment (nucleus and cytoplasm) (Ladanyi, 2000). Since then other types of chimeric oncoproteins involving ALK have been found such as, TPM (tropomyosin) 3 and 4 -ALK present in inflammatory myofibroblastic tumor (IMT) (Ladanyi, 2000). Fusion of the *ALK* gene to the clathrin heavy chain gene, CLTC, was found more recently in IMT (Bridge et al., 2001) and TRK-fused gene (TFG)-ALK was detected in other anaplastic lymphomas (Hernandez et al., 2002). Altogether, there have been reports of eight different ALK chimeric fusions that generate oncoproteins present in either IMT or anaplastic lymphomas (Bridge et al., 2001). In addition to this list it seems that ALK alone also has the potential to be involved in the genesis of highly malignant cancers of the brain such as glioblastoma multiforme (Powers et al., 2002). The full-length mammalian *ALK* has been shown to be expressed in the developing nervous system and therefore by virtue of its expression pattern is implicated in nervous system physiology (Morris et al., 1997).

Because the normal function of ALK remains unknown the downstream signalling molecules identified so far were all in cancer situations. The mitogenic effect of NPM-ALK has been attributed to the activation of the PLCγ pathway (Ladanyi, 2000). On the other hand the anti-apoptotic effect of the same fusion protein seems to utilise the PI3-kinase/Akt pathway as a result of the binding of the growth factor pleiotrophin (Ladanyi, 2000). This seems also to be the case in glioblastomas where the pleiotrophin-ALK signalling axis has been reported to have a rate -limiting factor in their growth (Powers et al., 2002). Stoica et al. (2002) has recently identified Midkine, a pleiotrophin homologue, as a growth, survival, and angiogenic factor during tumourigenesis that signals through the ALK receptor. The signalling transduction pathways involving Midkine-ALK were the anti-apoptotic PI3-kinase/Akt and the MAP-kinase pathways. Midkine was reported to bind to the ALK ligand binding domain with lower affinity than pleiotrophin in cultured cell lines (Stoica et al., 2002).

The orphan receptor LTK (Leukocyte Tyrosine Kinase) was the alternative candidate orthologue for the *shd* gene. Literature reports on LTK are scarce, perhaps because its clinical relevance is not as crucial as the *ALK* oncogene. LTK appears to utilise two main substrates; insulin receptor substrate 1 (IRS-1) and Shc. Shc mediates the interaction between Grb2 and Sos through tyrosine 862 at the carboxyl-terminal domain of LTK. IRS-1 phosphorylation, on the other hand, is done at tyrosine 485 at the juxtamembrane domain of LTK. Both of these tyrosines are located in the NPXY motifs that were identified as the consensus sequence of binding sites for the phosphotyrosine binding domains (PBT) of Shc and IRS-1. While the phosphorylation of both these tyrosines activates the Ras pathway generating mitogenic signals involved in cell proliferation, the phosphorylation of tyrosine 485, crucial for IRS-1 activation, generates a cell survival signal (Ueno et al., 1997; Ueno et al., 1996). In summary, both ALK and LTK have the potential to mediate cell proliferation and/or cell survival during disease situations but little is known about their normal function.

Other RTKs have been identified in zebrafish pigment cell proliferation and survival (Parichy et al., 2000b; Parichy et al., 1999). The zebrafish *c-kit* gene (encoded by the pigment pattern locus *sparse*) has a role in embryonic melanophore migration and survival. *Fms,* its closest known homologue (encoded by the pigment pattern locus *panther/pfeffer*) has been implicated in the maintenance (survival) of cells of the xanthophore lineage and therefore indirectly involved in the formation of the correct pattern of melanophore stripes in the adult zebrafish. Both these genes encode type III RTKs which belong to the platelet-derived growth factor receptor group (PDGFR) and are characterised for their extracellular Ig- like domains. On the other hand, both ALK and LTK belong to the type II RTKs or the insulin RTK subfamily. In this group we see very different structures of the receptors mature forms, although each is initially translated as a single polypeptide produced from a single gene locus. For example, the insulin and the insulin-like growth factor-1 receptors possess a disulfide-linked heterotetrameric α_2 β_2 structure, the MET and RON RTKs exist as αβ disulfide-linked heterodimers, and the mature TRK RTKs (TRK-a, -B, and -C) are present in the cell membrane as monomeric proteins (Barbacid, 1993; Gaudino et al., 1994; Fantl et al., 1993). Like the TRK RTKs, ALK and LTK are both monomeric proteins (Morris et al., 1997).

In this Chapter we utilise expression analysis on wild-type and *shd* mutant embryos and antisense technology using morpholino oligos to demonstrate molecularly that the *shd* gene corresponds to an *ALK* homologue as predicted in the previous Chapter by genomic gene finding analyses. We discuss whether *shd* is a common ancestor of *ALK* and *LTK* and we start focusing on the likely role of *shd* in the iridophore lineage.

4.2 Materials and Methods

4.2.1 Isolation of the complete ALK/LTK cDNA

4.2.1.1 Fish husbandry

Wild-type and mutant embryos were obtained from the zebrafish facility at the University of Bath, Department of Biology and Biochemistry. Crosses were set up as described in Westerfield (1995), eggs were collected from successful matings on the following morning. After rinsing, eggs were transferred to a Petri-dish containing embryo medium (5mM NaCl, 0.17mM KCl, 0.33mM $CaCl_2$, 0.33mM $MgSO_4$, 10^{-5}% methylene blue to suppress fungal growth). The embryos were raised at 28.5°C and staged according to Kimmel et al. (1995). Embryos to be live mounted or sorted were anesthetized in a 0.2% solution of 3-aminobenzoic acid ethyl ester ("Tricaine", Sigma) prior to analysis.

4.2.1.2 Primer design

Based on the results from Chapter 3 it was decided to design primers that amplified a region around 1Kb of the predicted ALK/LTK. This product would be cloned and sequenced and it would then be used to make an in situ riboprobe. The transcripts would reveal whether there was any neural crest expression pattern. According to our results from the cell autonomy study presented in Chapter 5 *shd* gene function was predicted to be cell autonomous and therefore we were expecting to see expression of the *shd* gene in iridoblasts or their precursors. When designing primers only exons that were given high probability by most programs included in the NIX package were considered. One important consideration was that the amplified region should not contain the tyrosine kinase catalytic

domain of the RTK gene. To avoid this conserved region the whole of the zebrafish predicted ALK/LTK cDNA sequence was compared to both human ALK and LTK genes using the BLAST application (www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html). After excluding these regions and finding high probability exons, primers were designed with the use of two programs. First, suitable primer pairs for PCR amplifications were suggested by Primer3 (http://waldo.wi.mit.edu/cgi-bin/primer/primer3.cgi/primer3_www.cgi). Primer pairs, each containing at least 1-2 GC clamps at the 3'end were then analysed for hairpins, dimers, crossdimers and melting temperature (Tm) with a second program, NetPrimer (http://www.premierbiosoft.com/netprimer/netprimer.html). Only primers with a rating of at least 80 (out of 100; NetPrimer) were selected. Primers were ordered from INVITROGEN Lifetechnologies with standard purification and 50 nmol synthesis scale. The primers obtained for the amplification of 1Kb fragment from the predicted zebrafish ALK/LTK were the following:

Forward shd1: GTGACAGTGCCTACCGAAACA

Reverse shd1: ATGATGGTGGACGAGCGAAT

4.2.1.3 Isolation of total RNA using Tri Reagent

Approximately 200 wild-type embryos at 24, 28 and 48.5 hpf were dechorionated using watchmaker forceps, killed by an overdose of the anesthetic Tricaine and homogenised thoroughly in 1ml Tri Reagent (Sigma). Tri Reagent is mixture of guanidine thiocyanate and phenol in a mono-phase solution that effectively dissolves DNA,RNA and protein on homogenisation or lysis of tissue samples. After adding chloroform and centrifuging, the mixture separates into three phases: an aqueous phase containing the RNA, the interphase containing DNA and an organic phase containing proteins. These components can then be isolated after separating the phases.

For the isolation of RNA the homogenates were spun at 13000rpm for 10 minutes at 4°C, the supernatant was removed to a fresh tube and the sample allowed to stand for 5 minutes at room temperature. 0.2ml chloroform was added, the tube shaken vigorously and left to stand for 2-15 minutes at room temperature. After centrifugation at 13000rpm for 15 minutes at 4°C the aqueous phase containing RNA was transferred to a fresh tube, 1/10 volume of isopropanol was added, the tube was inverted a few times and allowed to stand at room temperature for 5 minutes. The sample was spun again at 13000rpm for 10 minutes at 4°C and an equal volume of isopropanol added to the transferred supernatant. After the tube was inverted again and left to stand for 5-10 minutes at room temperature the precipitated total RNA was pelleted at 13000rpm for 10 minutes at 4°C. The supernatant was removed, the pellet washed with 1ml 75% ethanol by vortexing and the sample respun at 13000rpm for 5 minutes at 4°C. After removal of all the ethanol, the pellet was dried at room temperature for approximately 5 minutes and redissolved in 30μl of fresh milliQ water. The RNA was frozen at -80°C until necessary. The quality of the total RNA preparation was assessed by using a spectrophotometer (Beckman DU530) (OD 260nm/280nm should be around 2.0 for pure RNA). The concentration of the total RNA was calculated on the basis that an OD (260nm) =1 translates in a concentration of 40 μg/ml of RNA.

4.2.1.4 Reverse transcription using random hexamers

1-5μg of total RNA was combined with 250μl g random hexamers (Promega) and the volume made up to 12μl with fresh milliQ water. The sample was heated to 70°C for 10 minutes, chilled on ice, then 4μl of 5x reaction buffer (Gibco BRL), 2μl 0.1M DTT and 1μl 10mM dNTPs mix were added. After mixing the reaction, samples were incubated at 25°C for 2 minutes and 1 μl (200U) Superscript II (Gibco BRL) was added. Tubes were incubated for another 10 minutes at 25°C. All reactions were transferred to 42°C for 50

minutes followed by 15 minutes at 70°C to inactivate the reverse transcriptase. Then 2U RNAse H (Promega) was added and reactions incubated at 37°C for 20 minutes. The quality of the cDNA was primarily tested with primers known to work well on other previously synthesised cDNAs.

4.2.1.5 Polymerase Chain Reaction (PCR)

For all different PCR reactions, PCR conditions ($MgCl_2$, dNTP and primer concentrations) were optimised with a scheme of 12 reactions varying all parameters at the same time. The PCR reaction that worked for the primers forward and reverse shd 1 was composed of 1μl of cDNA, 4μl of 10x buffer (Thermoprime), 4μl Mgcl2 at 20mM, 4μl dNTPs mix (Gibco) at 2.5mM, 4μl of primers at 1μM, 0.8μl of Taq (Thermoprime), and 22.2 μl of MilliQ water. The PCR program that was used for this amplification was a touchdown programme with the first 10 cycles ramping from 65°C to 55°C in 10 seconds and then 30 cycles at 55°C. Denaturation step was at 92°C for 10 seconds and extension at 72°C for 4 minutes. This amplification was used for four cDNA time points; 24, 28, 48.5 and 63 hpf. The last time-point cDNA was kindly donated by S. Elworthy in the lab.

4.2.1.6 Cloning of PCR products directly into TOPO vectors

The PCR product obtained with the 48.5 hpf cDNA was cloned into TOPO TA vector (Invitrogen). The map is available at www.invitrogen.com. The cloning procedure followed the kit instructions. A cloning reaction was composed of 4μl of fresh PCR product, 1μl of salt solution and 1μl of vector. These 6μl were mixed gently and left at RT for 30 minutes. Then the cloning reaction was kept on ice until the one shot chemical transformation that followed. Bacteria were thawed on ice and 2μl of the cloning reaction were added to the bacteria tube and mixed briefly. These were incubated for 5-30 minutes on ice after which

the bacteria cells were heat shocked for 30 seconds at 42°C, without shaking, and returned to ice.

Meanwhile the SOC medium (2% Tryptone, 0.5% yeast extract, 10mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10mM$MgSO_4$, 20mM glucose) was thawed until room temperature was reached and selective Ampicillin LB plates were warmed at 37°C for 30 minutes. 250µl of SOC medium at RT were added to the tube containing the bacteria with the cloning reaction. This transformation tube was then shaked for 1 hour at 37°C at 200rpm. 20 and 50µl from each transformation were spread on two pre-warmed selective plates and incubated over night at 37°C. Colonies were streaked the next day and grown in 5ml of selective LB for further analysis. Minipreps were made with Wizard plus Kit (Promega) according to the Kit instructions. 50µl restriction digestions were done with EcoRI to release the inserts and check for their sizes on a 1% agarose gel.

4.2.1.7 Automated sequencing using BigDye terminator chemistry

The sequencing of plasmid inserts was carried out by Dr. Paul Jones (Sequencing Core Facility, University of Bath) on an ABI DNA sequencer using BigDye terminator chemistry. For sequencing, typically 300-500ng plasmids were combined with 10pmoles of primer or, alternatively, 3-10ng PCR product (approximately 500bp) with 5pmoles of primer in 6µl total volume. The resulting sequence files were viewed with the Chromas viewing programme. Sequences were then blasted against the NCBI databases to compare the cDNA clones with those predicted by the NIX package.

4.2.1.8 In situ hybridisation with zebrafish 1200 bp clone

4.2.1.8.1 Preparation of Dig labelled RNA in situ probe

After checking the orientation of the insert approximately 10μg of plasmid containing the 1200 bp clone of interest was linearised with 50U of restriction enzyme Spe I (Promega) in a total volume of 100μl. An aliquot of the reaction was loaded on a 1% agarose gel to check for complete digestion. The remaining sample was purified using the QIAquick PCR Purification Kit (Qiagen) and the concentration of the purified linearised plasmid estimated by comparison to a quantitative ladder. Probes were synthesised using the Dig RNA Labelling Kit (Boehringer/Roche) as described below.

In a 20μl reaction 1-2μg of linearised plasmid was combined with 1μl 20x Dig-NTP labelling mixture, 2μl 10x transcription buffer, 1μl RNAse inhibitor and 2μl RNA polymerase. T7 RNA polymerase was used to create an antisense RNA probe. After incubation at 37°C for 2 hours, 1μl aliquot was kept at –20°C. 1μl RNAse-free DNAse was added to the remaining reaction. The sample was incubated at 37°C for 15 minutes, then 15μl 5M Ammonium acetate (Ambion), 75μl absolute ethanol and 1.5μl seeDNA (Amersham) were added. The RNA was precipitated on ice briefly, spun at 13000 rpm for 15 minutes at 4°C, the pellet washed with 500μl 70% ethanol, air dried for 1 minute after carefully removing all traces of ethanol and redissolved in 20μl fresh milliQ water. Again, a 1μl aliquot was removed and compared to the one kept earlier before DNAse treatment on a 1% agarose gel. Before storing the RNA probe at –80°C 80μl formamide (Sigma) was added. As a starting point for in situ hybridisation 1μl of probe was diluted in 200μl of hybridisation mix (50ml hyb mix contains 25ml formamide, 12.5ml 20xSSC, 0.5ml 5mg/ml heparin, 0.5ml 50mg/ml tRNA, 0.25ml 20% Tween20, 0.46ml 1M citric acid and 10.7ml milliQ water).

4.2.1.8.2 In situ hybridisation on whole mount zebrafish embryos

Digoxigenin labelled RNA was detected with alkaline phosphatase conjugated anti-Dig antibody. To visualise labelled transcripts NBT/BCIP (Boehringer/Roche) was used as colour substrate in the alkaline phosphatase reaction.

Embryos from a cross between *shd* $^{ty82/+}$ carriers were used at the following stages: 22, 24, 30, 48,72 hpf. Staging was done according to Kimmel et al. (1995) and embryos were always dechorionated if they were older than 24hpf, anesthetised and fixed in 4% paraformaldehyde / PBS (0.8% NaCl, 0.02% KCl, 0.02M PO_4 pH7.3) for at least 24 hours at 4°C. Embryos older than 24hpf were incubated in a 0.003% phenyl-2-thiourea (PTU) solution to inhibit melanin sinthesis prior to fixation (Westerfield, 1995). This step prevented melanisation of melanophores which would obscure the detection of transcripts.

Embryos were dehydrated by washing 3x5 minutes in methanol, 1x10 minutes in methanol and then stored at –20°C for a minimum of 2 hours or until needed. Embryos were rehydrated by washing 5 minutes in 75% methanol/25% PBS, 5 minutes in 50% methanol/50% PBS, 5 minutes in 25% methanol/75% PBS and finally 4x5 minutes in PBT (0.1% Tween20 in PBS).

To improve the penetration of the probe, embryos between 18-somite and 22-somite stages were digested in a solution containing 10μg/ml proteinase K in PBT for 5 minutes, 24hpf embryos were treated for 15 minutes, 30hpf-48hpf embryos for 30 minutes and embryos up to 72hpf for 1 hour. Embryos were then refixed in 4% paraformaldehyde/ PBS for 20 minutes, washed 5x5 minutes in PBT and prehybridised at 65°C for 2-5 hours in hybridisation mix.

Embryos were hybridised in a minimum of 200µl of a 1:200 dilution of the probe and incubated at 65°C overnight. The probe was saved and stored at –20°C for re-use. To wash away non-specifically bound probe, embryos were first rinsed in hybridisation mix (HM), followed by 10 minute washes at 65°C with each of 75% HM/25% 2xSSC, 50% HM/50% 2xSSC, 25% HM/75% 2xSSC and 2xSSC in turn. Two high stringency washes at 65°C in 0.2xSSC for 30 minutes each were performed, followed by 10 minute washes at room temperature in each of 75% 0.2xSSC/ 25% PBT, 50% 0.2xSSC/50% PBT, 25% 0.2xSSC/75% PBT and PBT.

Blocking of the embryos was done at room temperature in PBT/2% sheep serum/2mg/ml BSA for 2-5 hours and incubation with the appropriate antibody (anti-Dig diluted 1:5000 in block solution) overnight at 4°C. The antiserum was discarded and the embryos rinsed once in PBT, then washed 6x15 minutes in PBT and 3x5 minutes in NBT/BCIP buffer (100mM TrisHCl pH 9.5, 50mM MgCl2, 100mM NaCl, 0.1% Tween20 in milliQ water). NBT/BCIP stain solution was freshly prepared by dissolving one NBT/BCIP tablet (Boehringer/Roche) in 10ml milliQ water. Embryos were stained in the dark for up to 48 hours and the reaction then stopped by washing the embryos in PBT several times. To improve the optical clarity of stained embryos they were transferred to 50% glycerol/50% PBS overnight or a few hours prior to inspection under an MZ12 dissecting microscope (Leica). For a more detailed analysis and for documentation the Eclipse E800 microscope (Nikon) with DIC optics was used.

4.2.1.9 RACE-PCR

RACE-PCR (Rapid Amplification of cDNA Ends-Polymerase Chain Reaction) was carried out using the Clontech SMART (Switching Mechanism At 5' end of RNA Transcript) RACE cDNA Amplification Kit. This method was used in order to amplify the zebrafish ALK/LTK-like cDNA ends. It was planned that by designing new primers in these ends, a Long

Distance PCR could then be performed to amplify the reminder of the cDNA. (primers are indicated in alignment 1of the Appendix).

First, 5'RACE cDNA and 3'RACE cDNA were synthesised by combining 1μg of total RNA from 48.5 hpf wild-type embryos with 1μl 5'-CDS primer or 3'-CDS primer respectively and 1μl SMART II (5'RACE cDNA reaction only). Supplied sterile water was added to a final volume of 5μl, the reactions incubated at 70°C for 2 minutes and cooled on ice for 2 minutes. After spinning the tubes briefly to collect the contents, 2μl 5x First-Strand buffer, 1μl 20mM DTT, 1μl 10mM dNTP mix and 1μl (200u) Superscript II (Gibco BRL) reverse transcriptase were added. Reactions were incubated at 42°C for 90 minutes, diluted with 100μl Tricine-EDTA buffer supplied with the kit and heat treated at 72°C for 7 minutes.

RACE PCR reactions were set up by combining 2.5μl 5'RACE or 3'RACE cDNA with 5μl 10x UPM (universal primer mix), 1μl 10μM 5GSP (5'gene specific primer) or 3GSP (3' gene specific primer), 5μl 10x Reaction Buffer, 1μl 10mM dNTP mix, 1μl 50x Advantage 2 Polymerase Mix (Clontech) and 34.5μl PCR-grade water. The gene specific primers that succeeded were designed based on the initial and terminal predicted exon sequences by the NIX package. Other, more internal, primers were used without success. It was therefore decided to try to amplify smaller fragments closer to the ends of the cDNA (for location see alignment 1 of the Appendix). The sequences from these gene specific primers were:

5GSP: CTGGTTGAAAGCGAGCGAGTTGGAGA

3GSP: CGTCTGTTTCTGCCCCGTCCTCTACC.

These were meeting the best requirements to use the SMART technology with Tm> 65°C, number of nucleotides between 23 and 28 and 50-70% GC content. A MJ Research PTC-DNA Engine PCR machine was used. Reactions were subjected to 5 cycles of 94°C for 10 seconds, 72°C for 3 minutes, then 5 cycles of 94°C for 10 seconds, 70°C for 10 seconds, 72°C for 3 minutes, then 30 cycles of 94°C for 10 seconds, 68°C for 10 seconds, 72°C for 3 minutes followed by a final extension step at 72°C for 10 minutes. 5μl of each reaction were run out on a 1% agarose/ethidium bromide gel and examined using an ultraviolet light source.

To test for product specificity nested 3' RACE PCR reactions were carried out by diluting 5μl of the primary PCR products into 245μl of Tricine-EDTA buffer. PCR reactions were set up as before, but replacing the 2.5μl cDNA with 5μl of the diluted primary PCR product and 1μl 10μM 3GSP with 1μl 10μM N3GSP. N3GSP sequence was as follows:

N3GSP: CCCCACCTCCATCTGCTGCACCCTCA

The nested PCR reaction was subjected to 25 cycles of 94°C for 10 seconds, 68°C for 10 seconds and 72°C for 3 minutes followed by a final extension of 72°C for 10 minutes. As before, 5μl of reaction were run out on a 1% agarose/ethidium bromide gel (TBE) and examined on an ultraviolet light source.

Cloning and sequencing of the RACE-PCR products was done in the same way as described for the 1200bp fragment (see section 4.2.1.6 and 4.2.1.7). Because the 3'RACE products were more than one, the stronger products were purified using the QIAquick Gel Extraction Kit (Qiagen) according to the Kit instructions.

4.2.1.10 Long Distance PCR

Primers were designed within the RACE fragments obtained in the previous section. One specific primer was designed in the 5'RACE fragment (5'LD) and different specific primers were designed in each of the three 3'RACE fragments obtained (3'LD1, 3'LD2, 3'LD3). All primers were chosen with the help of software Primer3 followed by Netprimer as mentioned in section 4.2.1.2. The primer sequences were the following:

5'LD: AGTGGACAGAGAGATTAGGCTA

3'LD1: GGGTGTTATCAGGTCAATTTGC

3'LD2: TGGTATTTCACGCACTTGGATA

3'LD3: GTTTTCACAGTGTTGACAGTAG

The system elected for performing Long Distance PCR (LDPCR) was the Herculase Enhanced DNA Polymerase (Stratagene). Other attempts were made using the Expand High Fidelity PCR System (Roche) without good results. Herculase enhanced DNA polymerase provided high fidelity amplification over more than 6Kb. It consists predominantly of *Pfu* DNA polymerase, combined with *PfuTurbo* PCR enhancing factor and a small amount of *Taq 2000* DNA polymerase. This Stratagene mixture allows amplification of genomic targets up to 37Kb while maintaining a lower error rate. PCR parameter optimisation was done according to the instruction manual the guidelines of which proved to be crucial.

The PCR reaction that was used successfully for the three required products was made of: 40.6μl of MilliQ water, 5μl of 10x Herculase buffer, 0.4μl of 25mM dNTPs mix, 1μl of 5' RACE ready cDNA, 2.5μl of 10μM primer mix and 0.5μl of (5U/μl) Herculase DNA

polymerase. The reagents were added on ice in this precise order. The 50µl reactions were then amplified in a MJ Research PTC-DNA Engine PCR machine using the following cycles: 2 minutes at 95°C, followed by 10 cycles of 30 seconds at 95°C, 30 seconds at 50°C and 5 minutes at 72°C. Another 20 cycles were then performed with 30 seconds at 95°C, 30 seconds at 50°C and 9 minutes at 72°C. PCR products were directly cloned into Zero Blunt TOPO PCR Cloning Kit (Invitrogen) as described in section 4.2.1.6. The reason for using this different vector was because the *Pfu* polymerases included in the Herculase solution make blunt end PCR products and therefore a suitable vector had to be used for cloning.

For the analysis of positive clones 12 colonies were grown overnight in LB containing 50µg/ml Kanamycin. Again plasmid minipreps using the Wizard Plus SV Miniprep DNA Purification System (Promega) were performed according to the Kit instructions. Sequencing of the three LD PCR products was done commercially by Oswel DNA Sequencing. Oswel synthesised all the sequencing primers on both strands of DNA. These are indicated in the wild-type sequences presented in the Appendix.

The three wild-type consensus sequences released by Oswel were compared to each other using the multiple sequence alignment package Clustalw available at EMBL-EBI (European Bioinformatics Institute (www.ebi.ac.uk/clustaw/). Open reading frames (ORFs) were analysed using the HGMP-RC package 'Getorf'. The amino acid sequences were also aligned using Clustalw. Identification of intron/exon boundaries was made by comparing these cDNA sequences with the PAC3 genomic sequence. The pairwise Blast (BLAST 2 sequences) was used for this purpose.

4.2.2 Morpholino gene knockdown of zebrafish *ALK/LTK- like* gene

Morpholinos are antisense oligos, typically 18 to 25 bases in length, designed to bind to a complementary sequence, referred to as the target sequence in a selected mRNA. This binding, if targeted to a translation start site, prevents translation of that specific mRNA (Ekker and Larson, 2001). As a consequence, the protein product coded by that particular mRNA is not made. Because morpholinos are stable, specific and non-toxic they are ideal for studies in embryos. Morpholinos can be easily delivered by microinjection into zebrafish 1-2 cell stage embryos (Ekker and Larson, 2001). In this work one morpholino was synthesised (Gene tools) against the beginning of the coding sequence for the mRNA encoded by the zebrafish ALK/LTK-like gene. The morpholino was ordered on line at the web site www.gene-tools.com and was designed by Gene tools free Oligo Design service. The first 25 bases of the coding sequence and as much 5'untranslated sequence as available were provided as well as the true AUG translational start site. The purchased morpholino had the following sequence:

AGTTTGTCGAGTAATATAATCCATG.

On its arrival the morpholino was resuspended in 1x Danieau buffer (58mMNaCl, 0.7mM KCl, 0.4 mM $MgSO_4$, 0.6 mM $Ca(NO_3)_2$, 5.0 mM HEPES pH 7.6) to make a stock concentration of 25μg/μl that was aliquoted and stored at 4°C. Microinjections were performed on 1-2 cell stage wild-type embryos. The procedure and equipment used were the same as described for the PAC genomic microinjections on Chapter 3 section 3.2.2.2. A range of decreasing dilutions of the stock solution was injected together with 10% phenol red. These resulted in the following increasing amounts per embryo: 1.15ng; 2.30ng; 4.60ng; 9.2ng; 18.4ng; and 36.8ng. Embryos were kept at 28.5°C until 72 hpf and were analysed under a MZ12 dissecting microscope (Leica) with incident light to observe

the number of iridophores. For Documentation a SPOT camera (Diagnostic instruments, inc.) was used.

4.2.3 Identification of the molecular lesions responsible for the different *shd* mutant phenotypes

RNA was extracted in the same way as described for the wild-type embryos in section 4.2.1.3. Around 100 homozygous mutants for the *shd* alleles *ty9*, *ty82* and *trd* were collected and kept alive until 72 hpf. At this stage they were easily identified and separated from their wild-type siblings and the RNA extraction took place. Filter tips were always used so that no contamination between mutants could take place.

The procedures used were exactly the same as those described for wild-type embryos, unless otherwise stated. 5'RACE ready cDNA (Clontech) had to be synthesised from each of the three mutant alleles so that the same LD PCR could work (see section 4.2.1.9) The primers used for the LD PCR with the three different mutant cDNAs were 5'LD combined with either 3'LD1 or 3'LD2 (see section 4.2.1.10 for sequences). Automated sequencing was performed directly on PCR products [typically 3-10ng PCR product (approximately 500bp) with 5pmoles of primer in 6μl total volume] after cleaning them with QIAquick PCR Purification Kit (Qiagen). An extra wash with 35% guanidine hydrochloride was included to remove primer dimers longer than 20bp.

Sequences were analysed using Bioedit software and the link to Clustalw to produce a consensus sequence for each mutant DNA. These were compared to the wild-type sequence to look for potential mutant lesions.

4.3 Results

4.3.1 Isolation of the complete zebrafish *ALK/LTK-like* gene

4.3.1.1 The 1200 bp fragment

We began by cloning a fragment suitable for asking whether *ALK/LTK* was expressed in the neural crest. We knew from cell autonomy studies (see Chapter 5) that the candidate gene for *shd* was expected to be expressed in the neural crest. Secondarily, the cDNA sequence would allow us to analyse the orthology of the RTK gene.

PCR using primers forward and reverse shd 1 was optimised (Figure 4.1) and conditions applied in the reaction in lane 11 were considered the best.

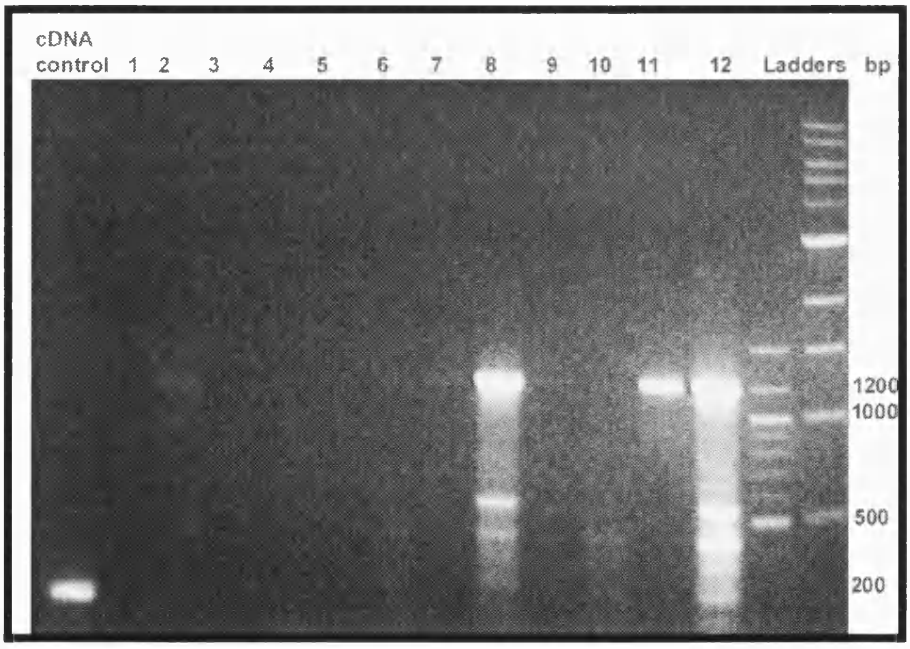


Figure 4.1. PCR optimisation with the primers forward and reverse shd1. The 1200bp fragment was clear in reaction 11. PCR conditions for reaction 11 were described in the Methods. A positive control (sox10 204bp fragment) was included in order to check the cDNA synthesis had been successful

Using these optimised conditions, PCR was performed on cDNA prepared from embryos of four different ages. This showed that the fragment was present at all stages but amplification was best at 48.5 and 63 hpf (Figure 4.2).

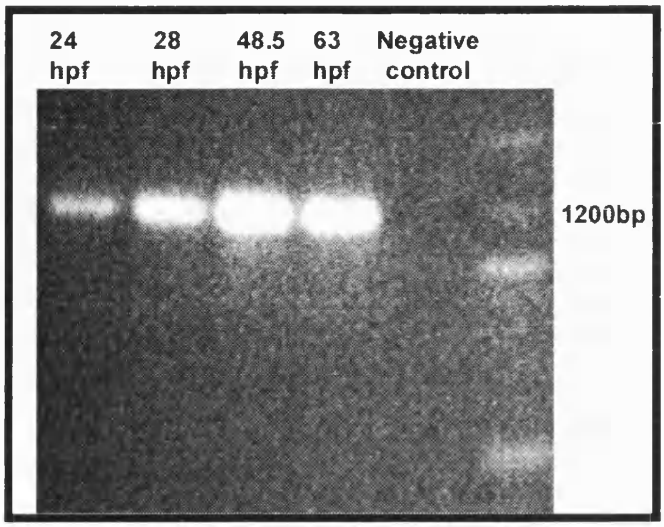


Figure 4.2. 1200bp fragment amplified from embryonic cDNA prepared from: 24, 28, 48.5, and 63hpf wild-type embryos. A negative control shows specificity of the amplified band.

Cloning of the 1200bp fragment amplified from the 48.5 hpf cDNA into a TOPO TA vector resulted in five positive inserts from 12 colonies (Figure 4.3).

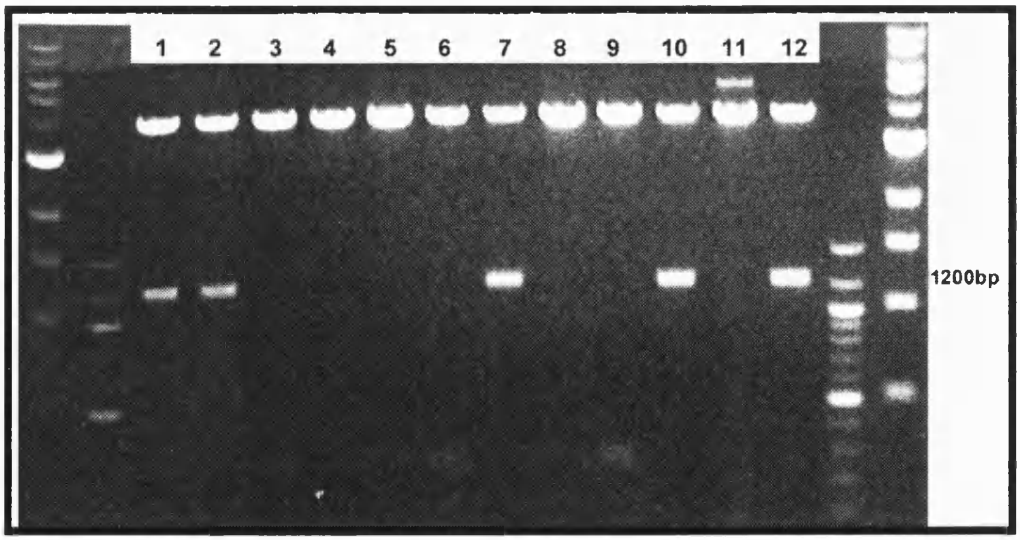


Figure 4.3. Result of EcoRI digestion of TOPO TA plasmid minipreps. Only five clones contained the expected 1200bp insert: clone number 1, 2, 7, 10 and 12.

Sequencing of the 1200bp fragment using the T7 sequencing primer generated sequence that allowed a translated BLAST search (Figure 4.4). This confirmed the high sequence similarity with ALK/LTK mammalian translated sequences.

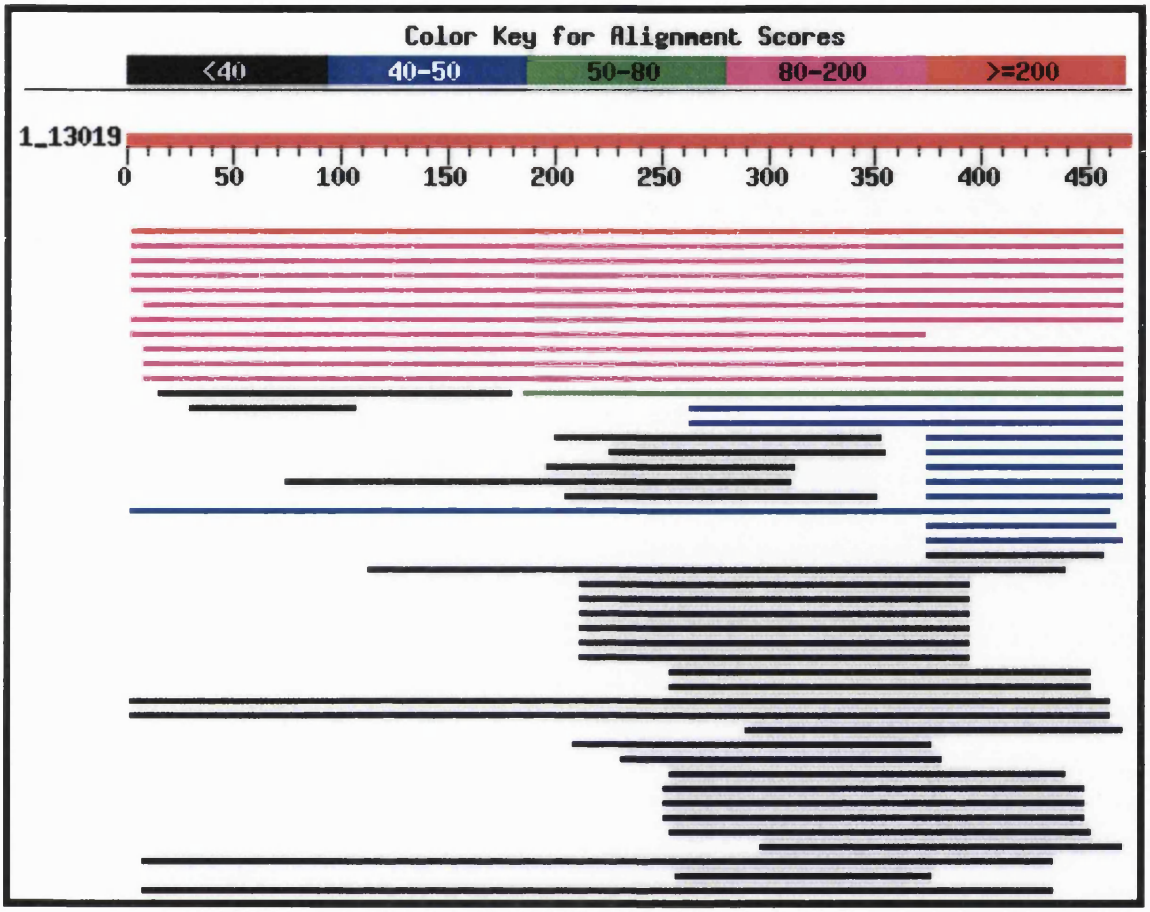


Figure 4.4. Translated BLAST result for the 5' 500bp of the 1200bp fragment. The best homology in red, not surprisingly, was with the zebrafish PAC3 genomic clone. The next ten best homologies in pink were with mouse ALK tyrosine kinase receptor precursor; human ALK tyrosine kinase receptor precursor; human ALK receptor, human LTK receptor precursor; human LTK receptor; mouse LTK receptor precursor; human non-catalytic LTK splice form; mouse LTK receptor; mouse protein tyrosine kinase; and rat LTK. The green trace is for rat ALK.

In conclusion, these results were encouraging and strongly supported the NIX predictions. A fragment of a zebrafish ALK/LTK-like gene had been cloned. The localisation of the full clone (1200bp) in the cDNA sequence is indicated in the Appendix (alignment 1).

4.3.1.2 Expression pattern of the 1200bp transcript

Confirmation of the clone identity led to the next question: what was the expression pattern conferred by a riboprobe made from this clone. To generate the riboprobe the plasmid containing the insert was first digested with Spe I. As expected this resulted in linearisation of the 5139bp clone (3939bp from the vector plus 1200bp from the insert) (Figure 4.5).

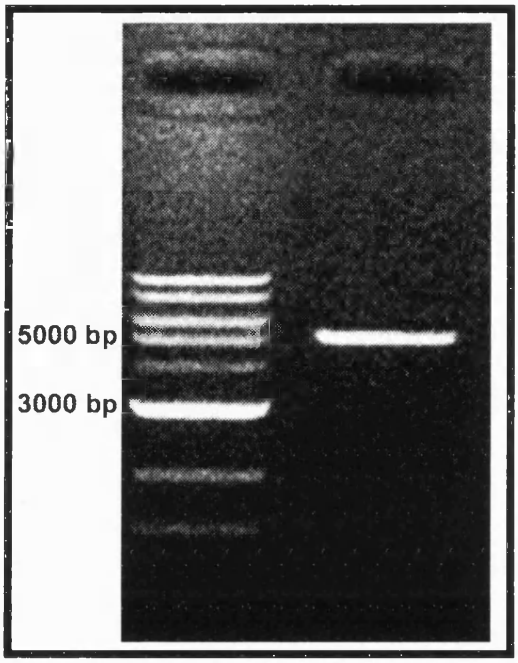


Figure 4.5. Making a riboprobe with the 1200bp fragment from the zebrafish ALK-LTK – like gene. 1% agarose gel showing the result of digesting TOPO TA vector (3939 bp) and the 1200 bp insert (= 5139 bp) with Spe I.

The 1200 bp riboprobe was used in a whole mount in situ hybridisation time course study. (see Figures 4.6 and 4.7). At 22 hpf it was possible to distinguish 25% of the embryos from the rest on the basis of the expression pattern in the head region (Figure 4.6 a-b). We therefore assumed these embryos were *shd* $^{-/-}$. Basically no expression or very reduced expression was detected in these *shd* mutant embryos when compared to their wild-type siblings. The expressing cells above the yolk sac and yolk elongation (Figure 4.6 c) were also almost absent in 25% of the embryos (Figure 4.6.d). Again we assume these were the *shd* $^{-/-}$ embryos resulting from an heterozygous cross. At this stage we did not assess in what tissues might the expressing cells be. A more detailed study, involving sectioning the

embryos is necessary. At 24 hpf these differences were also very obvious in the eyes (Figures 4.6 e and f). Small clusters of expression were easily seen in the iris of wild-types eyes resembling the iridophore later pattern whilst almost no staining was detected in *shd* mutants (25% of total number of embryos). At 30 hpf (Figure 4.7) transcripts were seen in the dorsal and ventral stripes of wild-types embryos at the right distance and number to be identified as iridoblasts (Figure 4.7 a, c). Again *shd* $^{-/-}$ mutants had virtually no expression (Figure 4.7 b, d). This pattern was maintained until 48 hpf when the eye expression in wild-types was perhaps most striking (Figure 4.7 e). At 72 hpf it looked as though the transcripts were being down regulated in the eyes and in the dorsal stripe (data not shown).

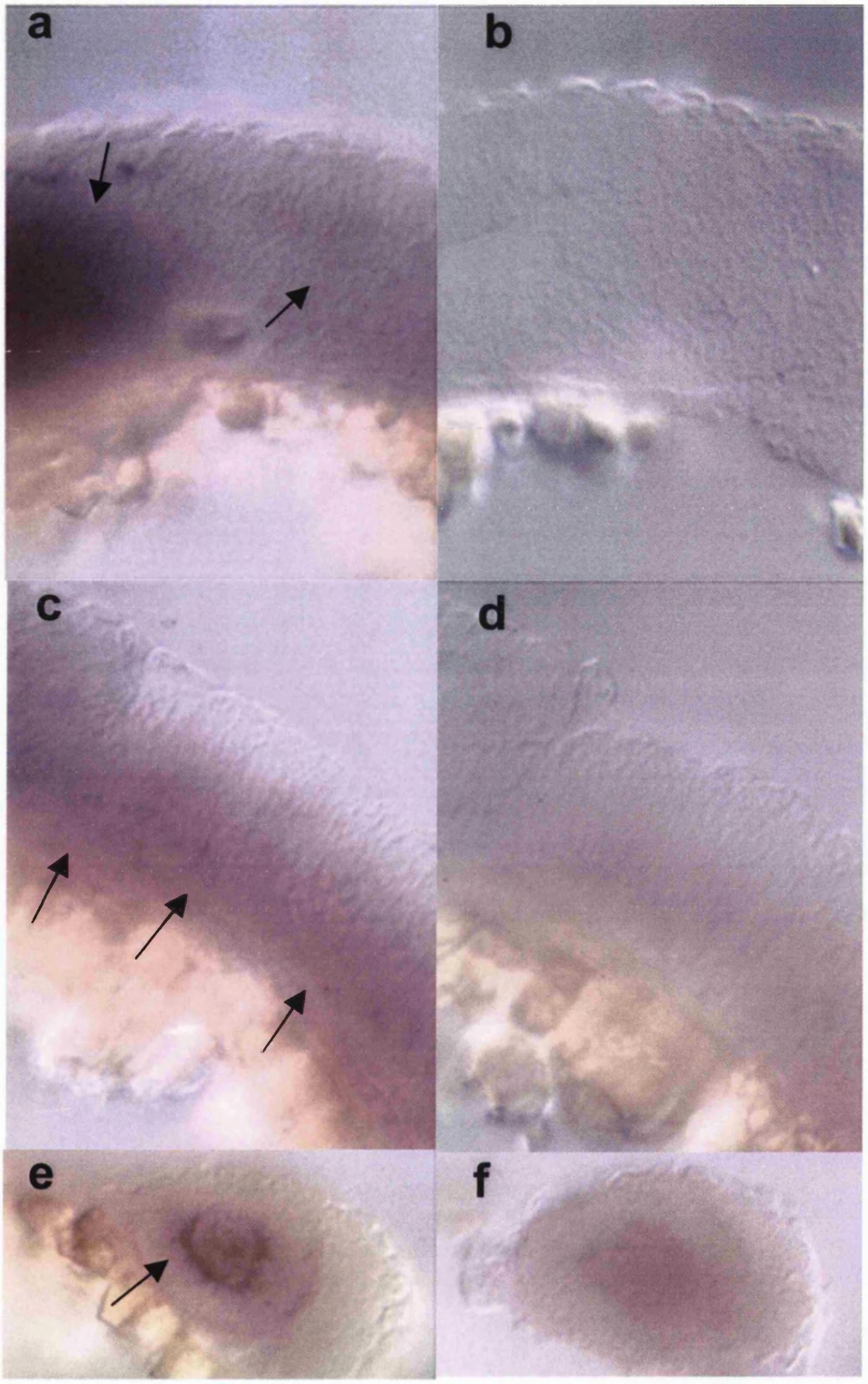


Figure 4.6. Whole mount in situ hybridisation with a zebrafish *ALK/LTK*-like riboprobe.**(a)** Expression at 22 hpf in the head of a wild-type embryo and **(b)** its absence in a *shd*^{ty82} mutant. **(c)** Expression at 22 hpf in a ventral stripe in the midbrain-hindbrain region in a wild-type and **(d)** its very reduced levels in a *shd*^{ty82} mutant. **(e)** Expression in the iris of a 24 hpf wild-type compared to **(f)** much reduced levels of expression in a *shd*^{ty82} mutant of the same age.

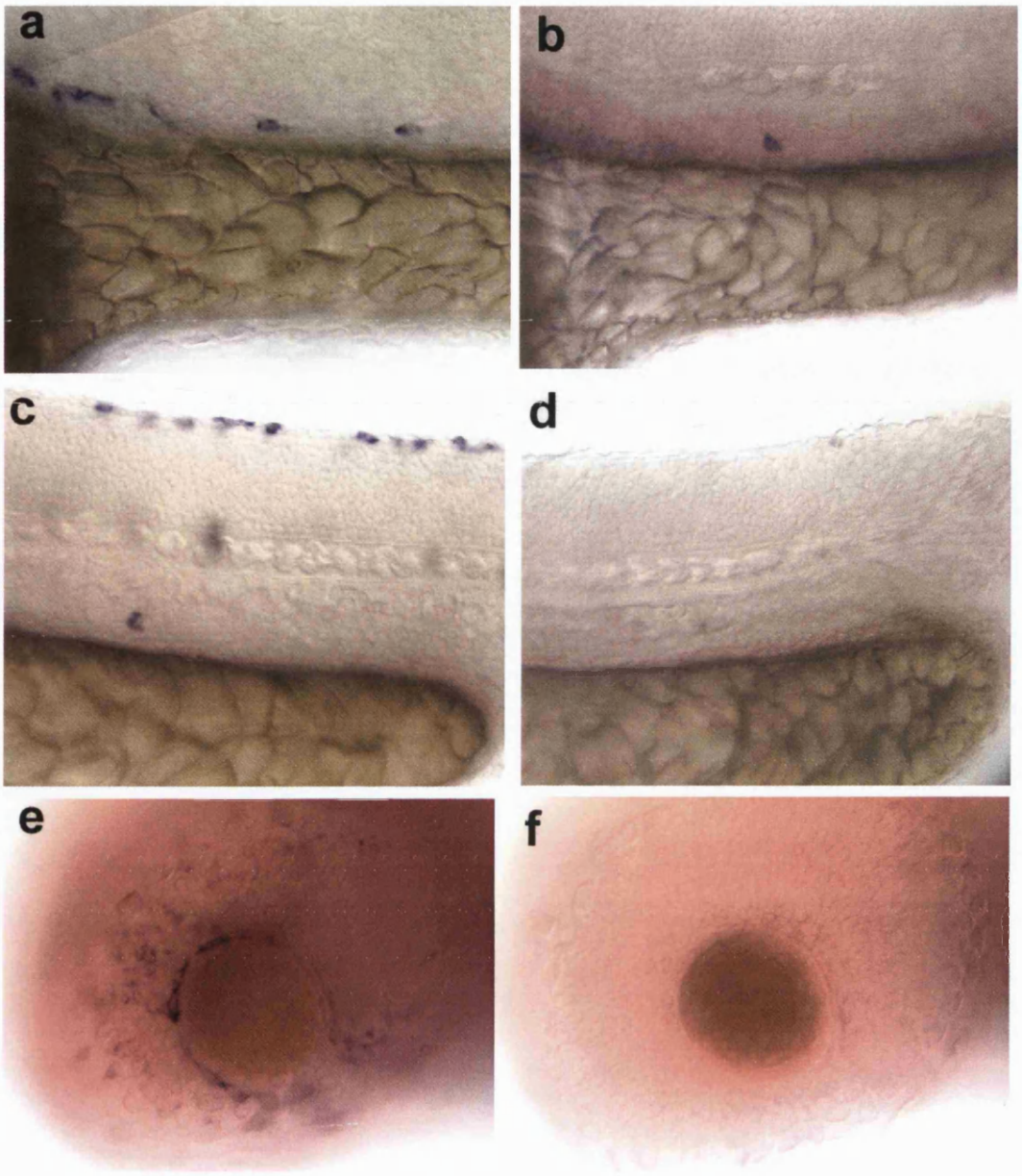


Figure 4.7. (a-f) Whole mount in situ hybridisation with a zebrafish *ALK/LTK* riboprobe. **(a)** Expression at 30 hpf is located in cells of the ventral stripe and **(c)** dorsal stripe of wild-type embryos. **(b,d)** These expression patterns are mostly absent in *shd* ^{ty82} mutants, respectively. **(e)** Expression in the eye of a 48 hpf wild-type embryo and **(f)** its absence in a *shd* ^{ty82} embryo of the same age.

4.3.1.3 The full zebrafish *ALK/LTK* cDNA

As the 1200bp sequence confirmed the identity of the expected RTK it was decided to clone the full length gene by performing RACE PCRs followed by Long distance-PCR. The RACE products generated were of different sizes: 700bp for the 5'RACE, and three different products of 800bp, 1250bp and 1500bp for the 3'RACE amplification (Figure 4.8 a). These 3'RACE products were confirmed with nested RACE PCR (Figure 4.8 b). Other much weaker products were also amplified in the nested RACE PCR.

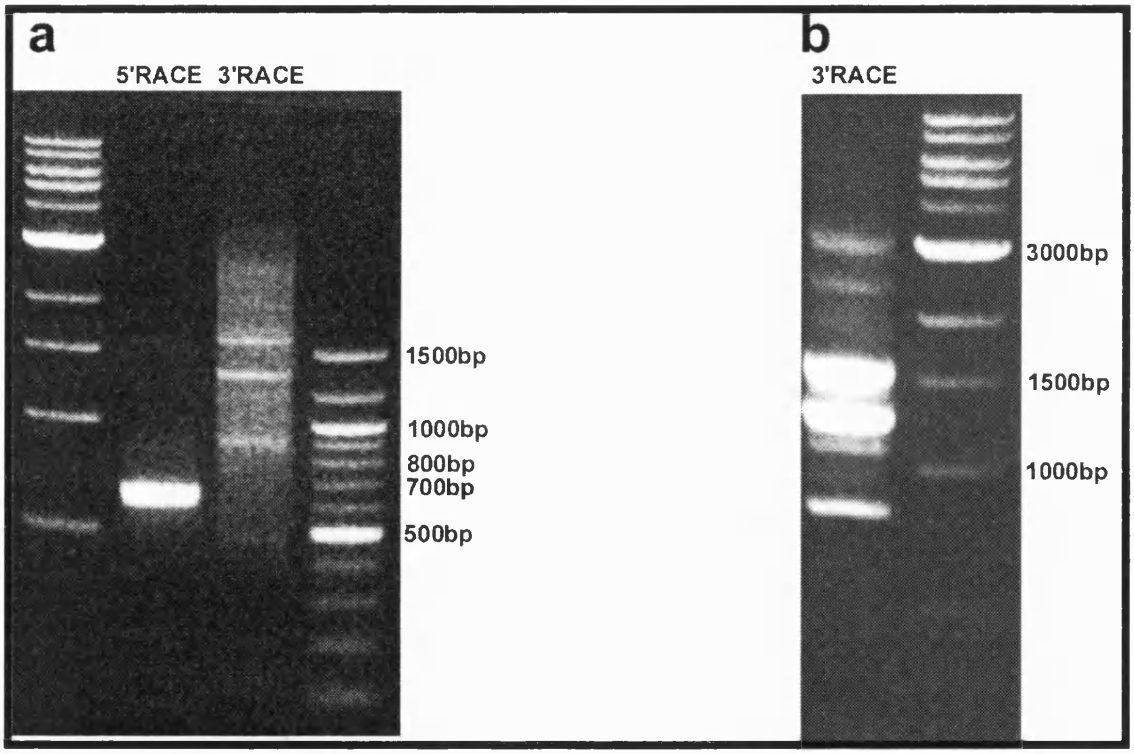


Figure 4.8. RACE PCR products (**a**) 5' and 3' RACE PCR products (**b**) 3'RACE products confirmed with nested PCR. The stronger bands in **b** correspond to the 3'RACE bands shown in **a**. The 5'RACE product is around 700 bp and the 3'RACE products are around 800, 1250 and 1500 bp long. Marker lanes show 1kb and 100bp ladders (Promega).

Analysis of the 5'RACE clones by restriction digestion with EcoRI showed that nearly all of them had the right insert around 700bp (Figure 4.9). The same procedure for the 3'RACE clones was not so successful, in particular for the 1500bp insert which was difficult to clone by gel purification (see Methods). However, a shotgun cloning of the PCR reaction was successful in cloning this fragment (Figure 4.10).

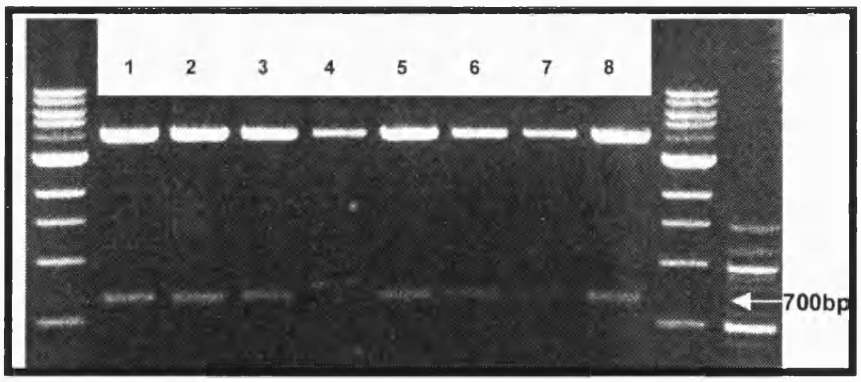


Figure 4.9. Cloning of 5'RACE products. Restriction digests of 5'RACE clones with EcoRI. Only clone number 4 did not contain the expected 700bp insert. Marker lanes show 1kb ladder on both ends of the gel and 100bp ladder on the right side of the gel.

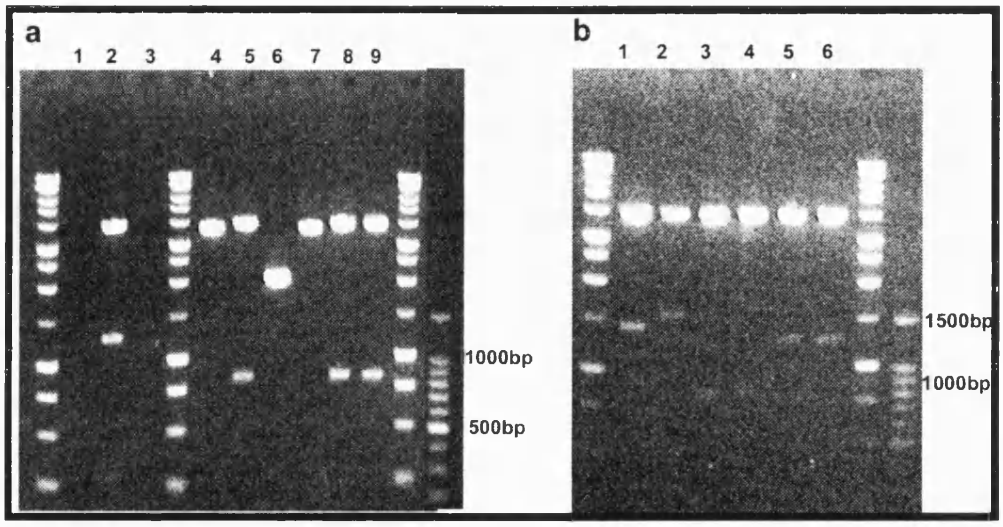


Figure 4.10. Cloning of the 3'RACE products. Restriction digest of the 3'RACE clones with EcoRI. (**a**) Clones from 1-3 resulted from the purification of a 1200bp band, only clone number 2 had the expected insert of 1200bp. Clones from 4-9 were generated from an isolated 800bp band. Clones number 5, 8, and 9 had the right size insert around 800bp.(**b**) Shotgun clones of the same PCR product. Clone number 2 had around a 1500bp insert. Clones number 3 and 4 had 800bp size inserts and clones 5 and 6 had 1200bp inserts. Marker lanes show 1kb ladders flanking the gels and a 100bp ladder on the right side of both gels

The purpose of the RACE PCR was to obtain sequence from both ends of the cDNA and use it to design new PCR primers that would be used in a LD-PCR aiming to amplify the transcript between the RACE products. As a result of the 3'RACE PCR three consistent products were obtained (Figure 4.10) leading to the design of three specific primers from each of the 3'RACE products and consequently to three LD-PCRs (all with the same 5' primer). A shorter one, around 5Kb, from the 3'primer originated with the shorter 3'RACE product, a medium one, around 5.5Kb, from the primer designed with the 1250bp 3'RACE

product, and a slightly larger one with the third primer originated from the 1500bp 3'RACE product (Figure 4.11). Cloning results in Figure 4.12.

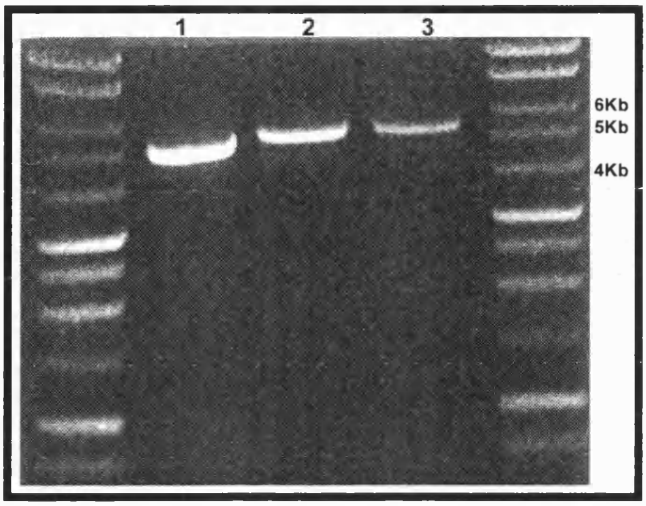


Figure 4.11. Long Distance PCR (LD-PCR) aiming to amplify near full-length cDNAs from the zebrafish ALK/LTK-like gene. Three products were obtained when using the three different specific 3'primers designed based on the three 3'RACE sequences Product 1 was around 5Kb and products 2 and 3 were around 5.5Kb. Marker lanes show a 1Kb ladder on each side of the gel.

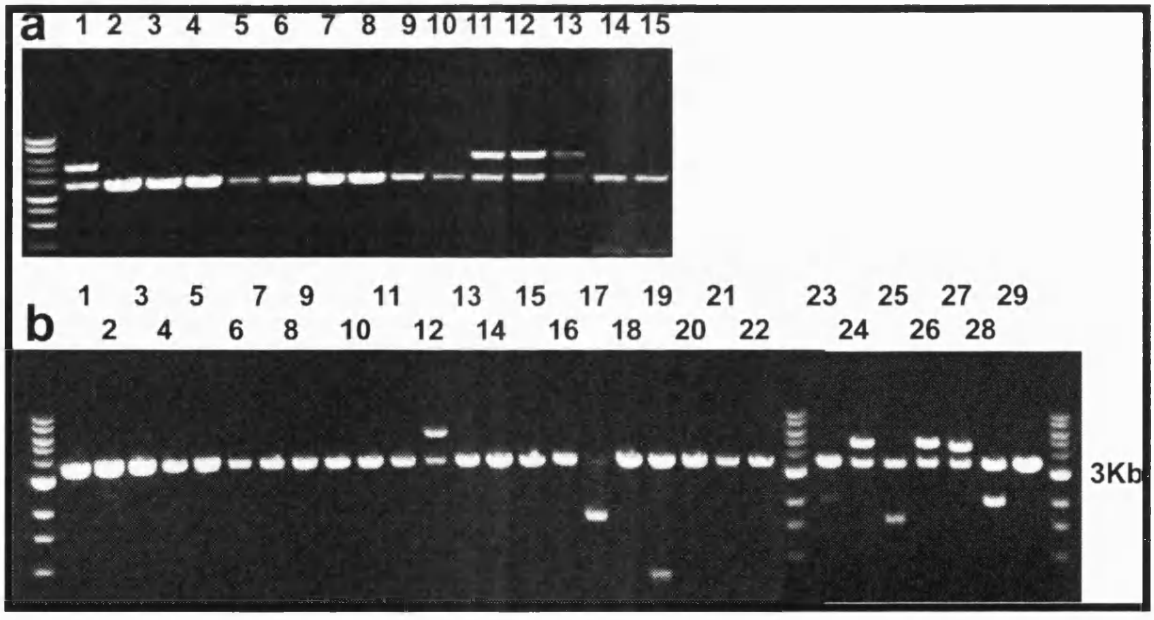


Figure 4.12. Cloning of the three LD-PCR products. Clones were digested with EcoRI to evaluate the insert sizes. **(a)** clone 1 (lane 1) had the right insert size around 5 Kb. Clone 2 (lanes 10-15) only lanes 11-13 had the right insert sizes around 5.5 Kb. Clone 3 (lanes 2-9) did not have any right insert sizes. **(b)** Clone 3 (lanes 1-22) only lane 12 had the right insert size. Clone 1 (lanes 23-29) only lane 27 shows the right insert size. Markers show 1Kb ladders.

Once all three bands were cloned and the DNA was fully sequenced it was possible to compare them to each other. The alignment produced for the three nucleotide sequences shows that they are very similar (see Appendix alignment 1). Clones 2 and 3 are identical but clone 1 has a 78bp insertion, beginning at position 2520bp, compared with clones 2 and 3. This results in a 26 amino acid insertion. The predicted ORF is otherwise, exactly the same. The start codon was assigned based on the morpholino results presented in the next section. The sequence shows two in-frame Methionin codons making the putative start site ambiguous. A morpholino designed to the first putative start site proved successful and therefore led us to believe that the ATG at position 83 was the right start codon. In this way, the zebrafish ALK/LTK-like gene open reading frame predicts a 1530 (clone 1) and 1504 (clones 2 and 3) amino acid protein. The sizes of the 3'UTRs from these three identical cDNAs are slightly different because clones 2 and 3 are truncated versions of clone 1 (see alignment 1 in the Appendix).

Comparison of the three cDNA sequences with the genomic sequence from PAC3 (Table 4.1 and Figure A1 in the Appendix) showed that the cDNA from clones 2 and 3 spread through the same 28 exons of genomic sequence (see Table 4.1). The only difference was that clone 2 was longer than clone 3 as already mentioned. In this way it was concluded that clones 2 and 3 probably represent the same transcript. In relation to clone 1 it was shown that the extra 78 nucleotides resulted in a larger exon 18 when compared to clones 2/3. Clone 1 also had an extra exon in the end separated by an intron of 32bp (see Table 4.1 and Figure A1). This last exon 29 did not contain any coding sequence but was the result of a longer 3'UTR unique to clone 1, absent in the other clones, perhaps due to truncation. Based on these results we believe we have cloned 2 different transcripts of the same gene: one longer one represented by clone 1 and a shorter version represented by clones 2 and 3.

Table 4.1. Genomic distribution of cDNA clones 1, 2 and 3

Exon number	Genomic sequence (bp)	CDNA sequence (bp)	Exon size (bp)	Intron size (bp)
1 (all clones)	4466-4579	*24-137	113	28197
2 (all clones)	32776-32870	136-230	94	2763
3 (all clones)	35633-35784	230-381	151	1890
4 (all clones)	37674-37901	379-606	227	3292
5 (all clones)	41193-41337	604-748	144	432
6 (all clones)	41769-41911	746-888	142	3662
7 (all clones)	45573-45723	888-1038	150	2987
8 (all clones)	48710-48812	1036-1137	101	639
9 (all clones)	49451-49614	1135-1299	164	2036
10 (all clones)	51650-51745	1299-1394	95	5472
11 (all clones)	57217-57293	1393-1469	76	481
12 (all clones)	57774-57953	1466-1645	179	4364
13 (all clones)	62317-62467	1645-1795	150	173
14 (all clones)	62640-62771	1796-1927	131	260
15 (all clones)	63031-63190	1927-2086	159	2625
16 (all clones)	65815-66002	2083-2270	187	2189
17 (all clones)	68191-68291	2266-2366	100	4068
18 (clone1)	72359-72705	2365-2711	346	5438
18 (clones2&3)	72359-72513	2365-2633	268	5630
19 (clone1)	78143-78330	2712-2899	187	1298
19 (clones2&3)		2634-2821	187	
20 (clone1)	79628-79720	2897-2989	92	82
20 (clones2&3)		2819-2911	92	

21 (clone1) 21 (clones2&3)	79802-79868	2988-3054 2910-2976	66 66	5645
22 (clone1) 22 (clones2&3)	85513-85644	3053-3184 2975-3106	131 131	865
23 (clone1) 23 (clone2&3)	86509-86607	3185-3283 3107-3205	98 98	170
24 (clone1) 24 (clones2&3)	86777-86873	3280-3376 3202-3298	96 96	78
25 (clone1) 25 (clones2&3)	86951-87057	3374-3479 3296-3401	105 105	194
26 (clone 1) 26 (clones2&3)	87251-87386	3477-3612 3399-3534	135 135	343
27 (clone1) 27 (clones2&3)	87729-87820	3613-3704 3535-3626	91 91	1238
28 (clone 1) 28 (clone2)	89058-90848 89058-90823	3701-5492 3623-5389	1791 1766	32
29 (clone1)	90880-90930	5564-5614	50	

*sequence 1-23 bp belongs to vector (see Alignment 1 in Appendix)

4.3.2 Morpholino gene knockdown of zebrafish *ALK/LTK-like* gene phenocopies the *shd* phenotype

Morpholinos are oligonucleotides of a specific morpholine backbone chemistry that when targeted to the translation start site result in efficient inhibition of protein translation (Ekker and Larson, 2001). The choice of the right start codon is therefore crucial for the morpholino to work. When designing the zebrafish *ALK/LTK–like* morpholino a careful search based on sequence comparisons with other organisms was performed in order to select the right start codon. The morpholino location in relation to the start codon can be found in the Appendix (alignment 1). Results from the morpholino injections are summarised in Table 4.2

Table 4.2 Summary results of the morpholino injections at different concentrations

Concentration (ng/embryo)	*No. injected embryos*	*% morphants*	*% abnormalities*	*% mortality*
1.15	140	4.2	10	–
2.30	178	1.2	9	9
4.60	163	2.3	16.5	22
9.2	260	65.5	5	30.7
18.4	255	82.7	8.6	9
36.8	38	84.8	27.2	13

When the data was analysed per batch of eggs injected as opposed to total number of eggs injected it was possible to show the variability associated with each batch. Such

variability could have resulted from the quality of the eggs, the skill of the injector on each particular batch , or even the quality of the needle being used. The results of treating the data this way are presented in Figure 4.13.

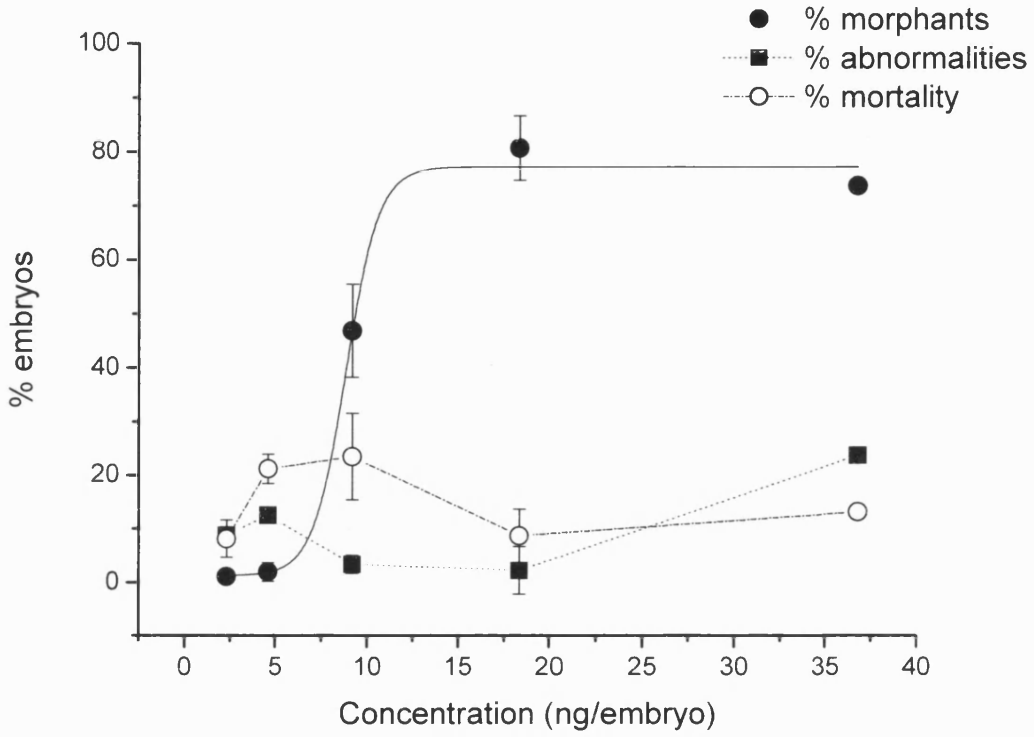


Figure 4.13. *ALK/LTK* morpholino injection phenocopied the *shd* phenotype. Graphic representation showing how the different concentrations of morpholino injected affected the percentage of abnormalities, mortality and successful morphants generated. A curve fit following a Boltzmann function [$y=(A_1-A_2)/1+e^{(x-x0)/dx}+A_2$] ($Chi^2=146.25874$ and $R^2=0.98889$) was applied to the % of morphants generated.

The analysis of these data led to two important conclusions: The first one being that the optimum concentration needed to phenocopy the *shd* mutant phenotype at 72 hpf was high (18.4 ng/embryo) and second that even such a high concentration had little associated non-specific abnormality or mortality, and was therefore considered ideal for this specific morpholino. The curve fitted according to the Bolzmann function indicated a concentration of 8.8 ng/embryo was needed to obtain half the number of morphants. This

model is typical of a concentration dependent event and shows that the number of morphants generated increases very fast (slope of 0.89327) with concentrations between 5 and 15 ng/embryo and that around 15 ng/embryo this number reaches a steady state meaning that a higher concentration will not result in more morphants being generated. The graphic also seems to indicate that the % of mortality and abnormalities are not concentration dependent. These parameters are probably more affected by the injection procedure and the damage it can cause to the eggs. Figure 4.14 shows the phenocopied wild-type embryos compared to uninjected wild-types.

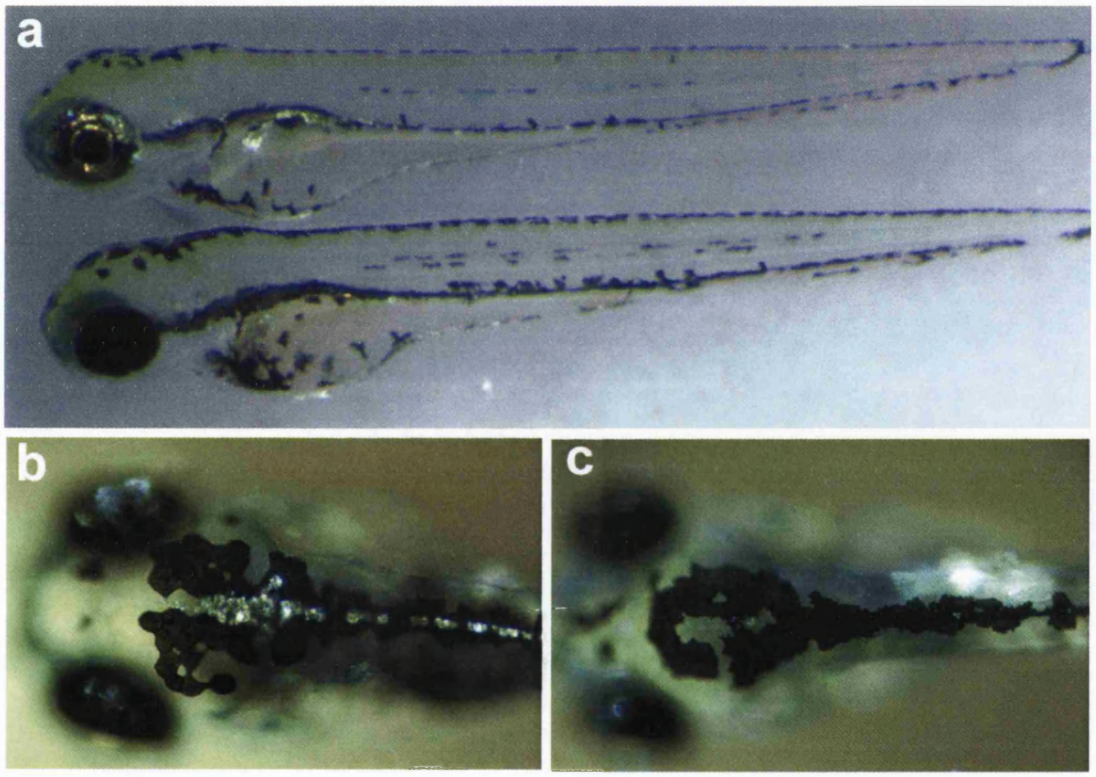


Figure 4.14. A morpholino designed to the zebrafish ALK/LTK-like gene phenocopies the *shd* phenotype. Injections were performed into 1-cell stage wild-type embryos. (**a**) Uninjected wild-type embryo (top) at 4 days post fertilization (dpf) and morphant of the same age (bottom). (**b**) Dorsal view of the head and dorsal stripe of a 4 dpf wild-type embryo and (**c**) of a morphant of the same age. The concentration of morpholino injected was 18.4ng/embryo. Phenocopied embryos showed no iridophores or a very reduced number of iridophores and were often indistinguishable from homozygotes for the strong *shd* mutant allele *ty82*.

4.3.3. Identification of the molecular lesions responsible for the different *shd* phenotypes

Conclusive evidence that a certain gene has been mutated is given by showing the altered sequence of mutated alleles of that gene. Therefore we attempted to amplify the cDNA from *shd* mutants *ty82, ty9* and *trd* by performing the same LD-PCR used for the wild-type embryos (see Materials and Methods). The three bands, one for each of the mutant alleles were consistent but not very bright which made direct sequencing difficult due to low concentration of DNA (Figure 4.15 shows an exceptional LD-PCR). Cloning of these products did not work in the available period of time probably due to poor amplification and DNA loss during band purification.

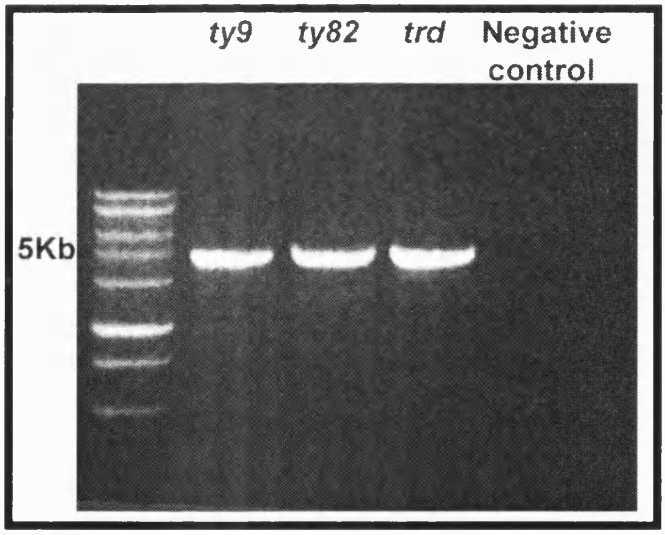


Figure 4.15. LD-PCR performed on *shd* mutants alleles *ty9*, *ty82* and *trd*. A negative control was included in the last lane to test for contamination. Marker lane shows a 1 Kb ladder.

Nevertheless, some sequence was obtained from the PCR product of the *ty9* mutant allele. Comparison of sequences from both strands with the correspondent wild-type region identified only one potential point mutation. A T to G nucleotide change would be predicted to result in replacing a Valine with a Glycine at the transmembrane domain residue 833. However, subsequent analysis of other wild-type clones showed that this occurred

frequently and was therefore a normal polymorphism and not due to the mutagenesis. This aminoacid replacement seems to be specific to zebrafish since this Val 833 is conserved between human and mouse (see alignment 4 in the Appendix).

Discussion

In this Chapter we have 1) confirmed that there is a zebrafish gene of the type II RTK subfamily in PAC3 that is similar to both mammalian ALK and LTK as predicted in Chapter 3; 2) shown that the zebrafish ALK/LTK homologue is expressed in neural crest; 3) identified its' genomic/cDNA structure and 4) provided evidence that it is indeed encoded at the *shd* locus.

Although the predictions for the genomic/cDNA structure given in Chapter 3 were very good they were not 100% correct. Most software included in NIX predicted a zebrafish gene with 20 exons that covered a PAC3 genomic region from around 32614 bp to 89061-90032bp. It was shown in the current Chapter that there were 29 exons and these spread from 4466bp to 90930bp of the PAC3 sequence.

Comparison of the two zebrafish ALK/LTK translated sequences and those available from other organisms in the SwissProt database showed that the zebrafish RTK predicted ORF is more identical to an ALK protein than to an LTK (E=0.0 for all blasts). The similarity to ALK proteins is found on all the domains over 1443 residues. LTKs do not have MAM domains and therefore the similarity to zebrafish RTK is based on the remaining domains spread over 759 residues.

The main region of conservation, conserved even in Drosophila ALK, corresponds to the tyrosine kinase catalytic domain between residues 865-1201 of the zebrafish clone 2/3. Based on this region the zebrafish RTK shows 10% more similarity to human LTK than to human ALK. This catalytic region included the consensus ATP binding sequence GxGxxG followed by the AxK residues downstream (for localisation of conserved domains see

alignment 2 in the Appendix). This domain was also confirmed by the NCBI Conserved Domain Search Database (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) with an E value of E=4e-94.

Other conserved domains identified were the two MAM domains at residues 53-218 and 277-431(E values of E=6e-06 and E=1e-20, respectively). MAM domains comprise about 160 amino acids found in extracellular domains of many receptors likely to have an adhesive function relevant in cell/cell interactions (Beckmann and Bork, 1993; Jiang et al., 1993). The MAM domain is named after meprins, A-5 protein and receptor protein tyrosine phosphatase mu. Such a domain is found in mammalian ALK and Drosophila ALK but not in LTK (Loren et al., 2001).

A domain that is found in both proteins ALK and LTK and is also in this zebrafish RTK is the LDLa domain (Low Density Lipoprotein Receptor Domain class A) which is known to play a central role in mammalian cholesterol metabolism (Daly et al., 1995; Fass et al., 1997). In zebrafish it is located between residues 240 and 264 (E=7e-4) just before the second MAM domain. A schematic diagram clarifies these structural similarities between zebrafish ALK/LTK, Drosophila ALK, human ALK and human LTK (Figure 4.16). An hydrophobic transmembrane domain comprising 26 residues was identified by comparing the zebrafish sequences to mouse and human ALKs (Iwahara et al., 1997).

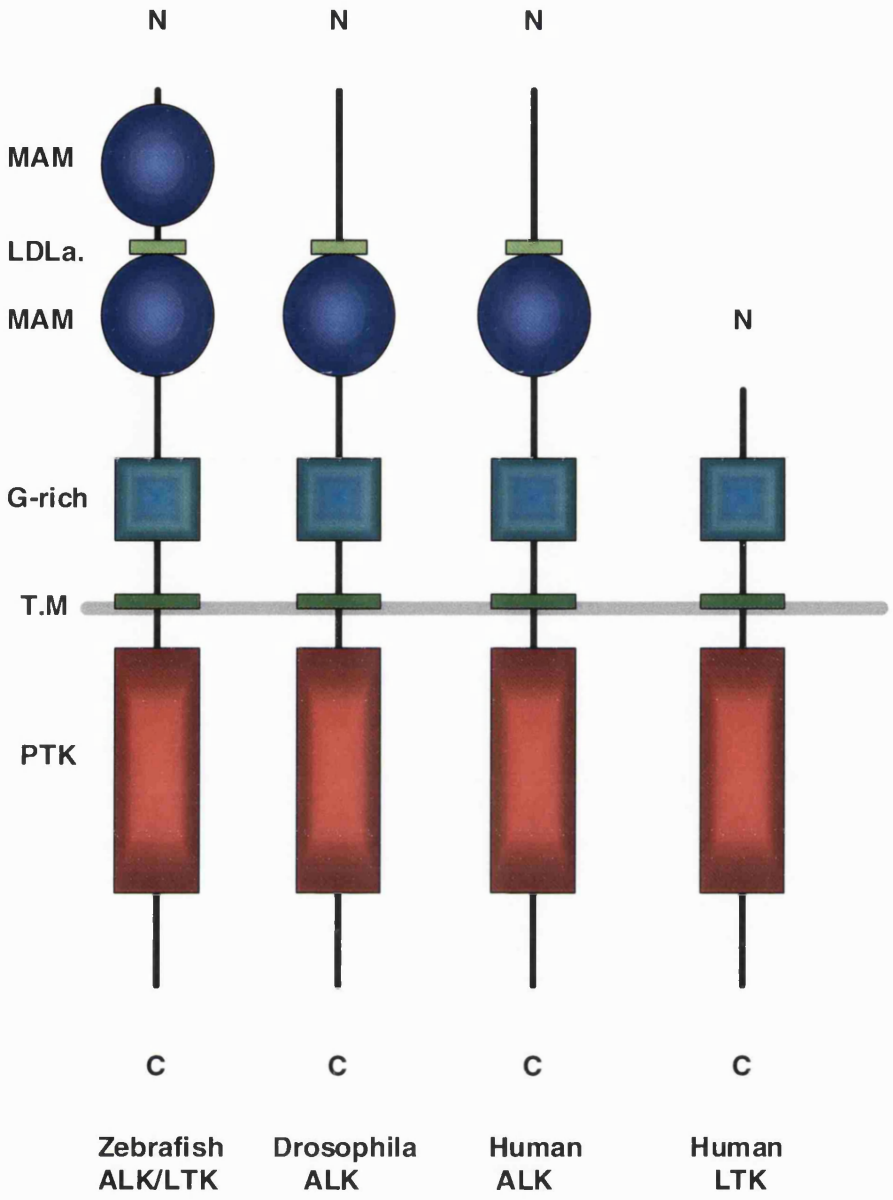


Figure 4.16. Structural similarity between zebrafish ALK/LTK, Drosophila ALK, human ALK and human LTK. Structural conservation includes the presence of an intracellular Protein Tyrosine Kinase (PTK) domain, a transmembrane domain (T.M.) and a glycine rich extracellular domain (G-rich). ALK proteins also share MAM domains that are not present in LTK. A low density lipoprotein class A (LDLa) domain is shared between zebrafish and the ALK proteins (adapted from Loren et al., 2001).

Based on the presence of MAM domains in the zebrafish protein it is suggested that the identity of zebrafish RTK is an ALK orthologue as opposed to a LTK orthologue. This argument was also given by Loren et al. (2001) when they named Drosophila ALK as such. Translation of zebrafish *shd/ALK*, as we shall refer to it from now on, shows two MAM domains similar to those found in mouse ALK and unlike human ALK where we can only find one such domain (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

The question remains on whether the zebrafish *shd/ALK* identified in this work is the true ancestor of mammalian *ALK* and perhaps also of *LTK*. This and other questions such as whether zebrafish *shd /ALK* is closer to human, mouse or Drosophila *ALKs* were addressed by a comprehensive phylogenetic study done in collaboration with Prof. Laurence Hurst (University of Bath, Dep. Biology and Biochemistry).

Using the nucleotide and translated sequences for both clones (with and without the 78 nucleotides) he submitted them for comparison with homologous sequences in a phylogenetic analysis, using the following seven sequences: AF236106 (Drosophila melanogaster ALK), U62540 (Human ALK), NM_007439 (Mouse Alk), AB073169 (partial rat Alk), NM_002344 (Human LTK), M90470 (Mouse Ltk) and, XM_230479 (rat Ltk). Results are presented in Figure 4.17. A very conservative approach was taken by restricting the analysis to the more conserved regions, the tyrosine kinase domains.

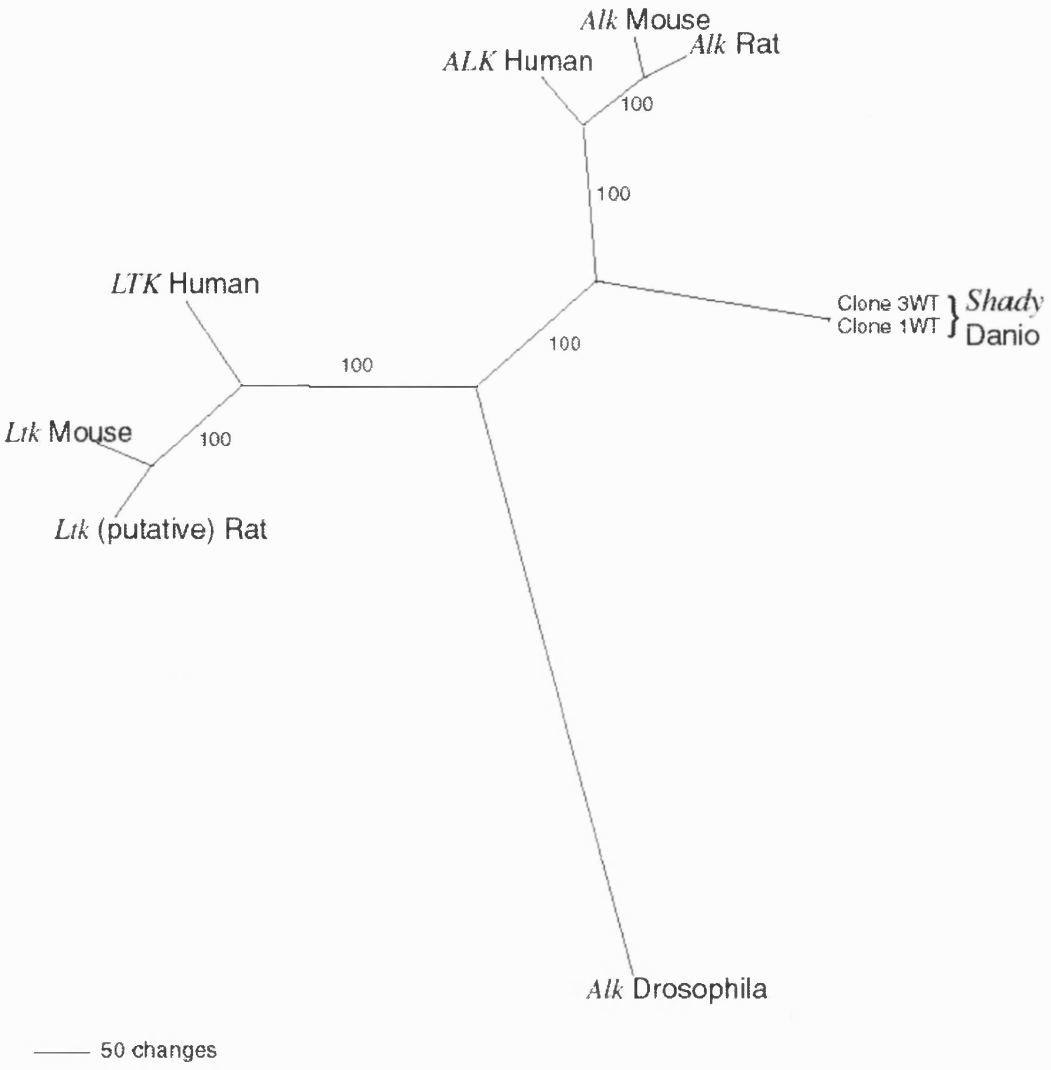


Figure 4.17. Phylogenetic analysis of the zebrafish RTK. Nucleotide and translated sequences from the zebrafish (*Danio rerio)* protein tyrosine kinase domain (PTK) were compared to the same domains from sequences AF236106 (Drosophila melanogaster ALK), U62540 (Human ALK), NM_007439 (Mouse Alk), AB073169 (partial rat ALK), NM_002344 (Human LTK), M90470 (Mouse Ltk), and XM_230479 (rat LTK). The coding sequences were extracted by reference to the annotations in the GenBank files using gbparse (http://sunflower.bio.indiana.edu/~wfischer/Perl_Scripts/). Alignment was performed using ClustalX on the translated sequences (Thompson et al. 1997). The nucleotide alignment was reconstructed from the protein alignments using AA2NUC.tcl (a tcl script written by Laurence Hurst, available on request). The nucleotide alignment was filtered by Gblocks (Castresana 2000) under the codon mode in default settings to eliminate ambiguous alignments. The phylogeny was estimated using a Bayesian method (Huelsenbeck & Ronquist 2001). One million generations were simulated resulting in 10,000 trees. The general time reversible model was assumed, with site specific rates with sites partitioned by codon. The last 5000 of the 10000 all have approximately the same likelihood and represent the 5000 most likely topologies. From these a 50% majority rule consensus tree was reconstructed. The figures on the branches in the Figure represent the frequency of occurrence of this topology within the set of 5000.

The phylogeny strongly supports that zebrafish clones are ALK homologues. Indeed looking at the protein alignment, we looked at two figures: 1) how many times we find both clones the same as both ALks but different from both Ltks; 2) how many times we find both clones the same as both Ltks but different from both Alks. The answer to the former question is 37 and to the latter is just 6. Therefore our clones are much more similar (at least in this PTK region) to ALK. We can thus confirm that this zebrafish RTK is an orthologue of mammalian *ALK*s.

So far both the mapping and the molecular data supported the hypothesis that the gene identified as zebrafish ALK was encoded by the *shd* locus. As an independent experiment to further test this hypothesis we injected a specific morpholino targeting the zebrafish ALK-like gene. It was predicted that by blocking the translation of this protein in wild-type embryos these would phenocopy the *shd* mutants and would therefore lack the majority of the iridophores.

With the morpholino we observed a high rate of phenocopied embryos. Although high levels of morpholino were used (18.4ng/embryo) abnormalities remained low suggesting a specific targeting. We believe the requirement for a high dose can be explained by the relatively late function of *shd*. Thus, most mutant phenocopy studies have used much lower concentrations of morpholino because genes functioning earlier in development were being targeted (e.g. Karlen and Rebagliati, 2001). Others have also injected high doses (15ng) with good results as late as 4 dpf (Miller and Kimmel, 2001). Perhaps the best comparison with the results presented here is the morpholino phenocopy of *colourless,* another pigment mutant that lacks all three types of pigment cells including iridophores (Dutton et al., 2001). These authors found that they could only achieve strong phenocopies with high doses of morpholino (optimal dose 16.5ng) (Dutton et al., 2001). Unfortunately this scoring was based on melanophores and did not include the other pigment cells. Nevertheless, if they needed such a high concentration for scoring a

pigment phenotype at 27 hpf then it is understandable that for a 72hpf phenotype such as *shd* 18.4ng might be necessary.

Because genomic rescue (Chapter 3), mRNA expression studies and morpholino knock down studies were all successfully performed a cDNA rescue approach was not undertaken. Supporting this decision was data in the literature suggesting that very long cDNAs (6055bp) may present a problem for heat-shock rescue experiments (Parsons et al., 2002). Using a non-endogenous promoter also presented a problem in the past (Lyons et al., 2002). As a last confirmation that the *shd* locus is encoded by the zebrafish ALK-like gene we intended to identify the mutant lesions but could not do these experiments in the available period of time. However these last experiments will be performed in the near future for the purpose of publication.

The most interesting question that emerges from the results presented in the current Chapter concern the relationship between the gene identified here and its putative mammalian homologues. Whilst the mammalian *ALK* is expressed in both the embryonic and neonatal CNS and PNS (Iwahara et al., 1997), the human LTK gene product was identified mostly in placenta and hematopoietic cell lines (Kozutsumi et al., 1993). In addition, mouse *LTK* mRNAs were detected in adult, but not embryonic brain, and lymphoid cells (Snijders et al., 1993). Zebrafish *ALK* showed specific expression in the embryonic iridophore lineage and, at around 22 hpf, in the brain as well as in a ventral elongated region above the yolk sac and yolk extension. Consequently, the only similarity in the expression patterns is in embryonic CNS between mammalian *ALK* and zebrafish *ALK*. On the other hand, the region of expression above the yolk needs to be properly identified because it could be part of the gut or the excretory system. As mammalian *ALK* transcripts have been reported in the enteric nervous system (Iwahara et al., 1997) and mamalian *LTK* in the kidney (Ben-Neriah and Bauskin, 1988) we need a more careful

determination of the exact tissue distribution of the zebrafish transcripts to permit accurate comparison with mammalian genes.

It has been shown that the growth factor pleiotrophin (PTN) and its homologue Midkine (MK) are both ligands for the mammalian ALK promoting growth, survival, and acting as angiogenic factors during tumorigenesis (Stoica et al., 2001; Stoica et al., 2002). Blocking of the ALK receptor with an ALK monoclonal antibody targeted to the ligand binding domain, within one of the mouse MAM domains, showed that MK and PTN could no longer bind to ALK. Furthermore, MK or PTN induced colony formation of SW-13 cells was inhibited supporting the idea that these growth factors are true functional ligands for the ALK receptor (Stoica et al., 2002).

Two zebrafish heparin binding growth factors, *midkines a* and *b*, (*mdka* and *mdkb)* have been identified as well as a EST for zebrafish pleiotrophin (*ptn*) (Winkler and Moon, 2001; Winkler et al., 2003). *mdkb* was previously known as *mdk2* (Winkler and Moon, 2001) and it has been shown that *mdkb* participates in posterior neurogenesis and that its ectopic expression enhances neural crest fates at the lateral edges of the caudal neural plate. On the other hand, *mdka* is involved in floor plate development (Winkler et al., 2003). So far there are no studies done on zebrafish *ptn* only on its phylogenetic relationship to both *midkines* that showed the separation between these genes arose before the divergence of the fish and tetrapod lineages at least 450 million years ago. Thus, it was concluded that *mdka* and *mdkb* were not fish versions of *PTN* (Winkler et al., 2003).

In contrast to the ubiquitous expression of mammalian *MK*, zebrafish *mdka* and *mdkb* are expressed in restricted, mostly non-overlapping, areas during embryonic development. Thus, *mdka* is expressed in the paraxial mesoderm and somites, central neural tube and brain whereas *mdkb* is expressed in regions of the dorsal CNS (Winkler et al., 2003).

From both mRNA expression and overexpression studies (Winkler and Moon, 2001) *mdkb* is the most likely candidate ligand to zebrafish shd /ALK. In order to test *mdkb* candidacy it would be of major interest to look at *mdkb* later expression pattern and investigate whether expression in the neural crest cell environment persists and remains active alongside with *shd /ALK*. If appropriate expression is observed it could be speculated that perhaps the same anti-apoptotic pathway (IP3-kinase/Akt) activated by the mammalian Pleiotrophin/Midkine binding to ALK (Stoica et al., 2002) is responsible for the survival of iridophore precursors. Then in the absence of *shd /ALK* function iridoblasts die.The last hypothesis would point to *shd/ALK* having a role in the survival of iridophores. Although cell death was not tested for, the in situ time course did not reveal expression in *shd* $^{-/-}$ mutants at any tested time point. Perhaps because *shd/ALK* function is needed to maintain its own expression in a positive feedback loop. Thus, iridoblasts might never develop at all in a *shd*$^{-/-}$ embryo, suggesting an earlier role for *shd/ALK* conceivably in specification. Alternatively, iridoblasts might be specified but their fate is not maintained.

Deeper understanding into the potential role of *shd/ALK* in the pigment cell lineage will be given in the next Chapters after presenting the expression pattern of different genes in *shd* mutants and wild-type embryos.

Chapter 5

Investigating the *shady* phenotype

5.1 Introduction

The *shd* complementation group included 12 mutations that formed a clear phenotypic series (*tc205; te295; te300; tf238c; th219; ti263c; tj229e; tm46a; tp218; ty9; ty70* and *ty82*) (Kelsh et al., 1996). Mutants in the strongest alleles (e.g. *ty82* and *te295)* have no more than a few iridophores if any, with remaining iridophores usually on the interior side of the eye only. Mutants for the weaker alleles (e.g. *ty70* and *th219*) show an approximately 50% reduction of iridophores in the lateral patches but a much weaker reduction in iridophore number elsewhere (see Figure 5.1 for location of body stripes and patches).

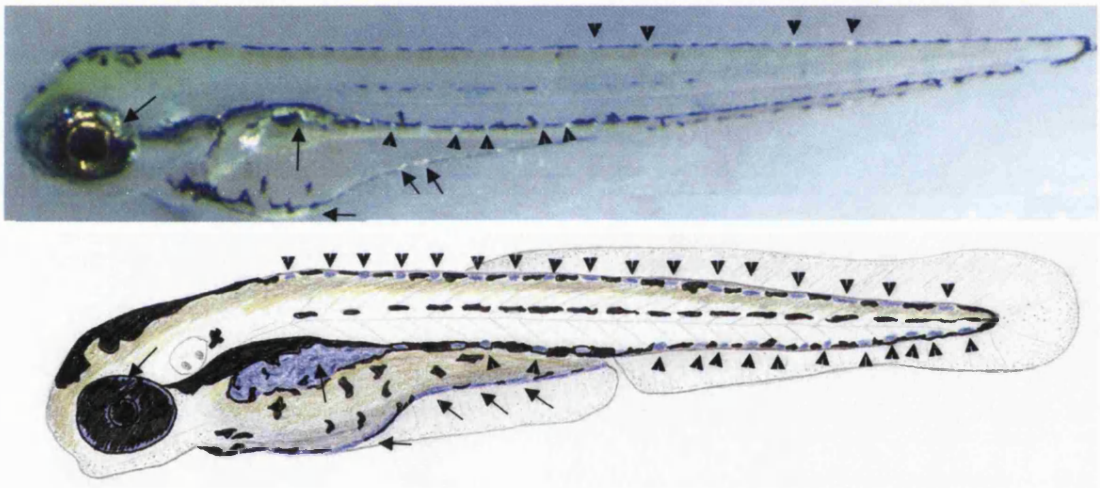


Figure 5.1. Wild-type Iridophore body pattern. Live embryo (5 dpf) photographed with incident light (top) and schematic drawing (bottom) for better identification of iridophore cells in blue. The yellow cast is given by xanthophores. Dorsal stripe with black melanophores and shiny (or blue) iridophores (down facing arrowheads). Shiny (or blue) patches of iridophores are visible in the eye and in the lateral patch as well as in the yolk sac stripe (arrows). Iridophores are also present in the ventral stripe (arrowheads facing up) alternating with melanophores.

All the other intermediate phenotypes show some iridophores in the eyes and lateral patches but relatively few in the yolk sac stripe.Thus, *shd* was classified as having one

pigment cell type reduced, essentially absent in strong alleles (Kelsh et al., 1996). Some weak alleles are homozygous viable and develop into adults with usually no obvious adult phenotype. Only one allele (*shd* ^{te295}) shows an adult phenotype; homozygotes show a reduction of both iridophores and melanophores in the adult and for this reason it is possible that its phenotype is due to a separate, closely linked mutation (Haffter et al., 1996b).

The *shd* phenotype fulfilled the criteria for mutations causing defects in chromatophore specification. These were expected to cause very severe pigment phenotypes, affecting one or more chromatophore types, and for strong alleles to show a total absence of that cell-type(s) (Kelsh et al., 1996).

With no more available information on the *shd* phenotype at the beginning of this project. I began to investigate the general aspects of this mutant phenotype alongside the main objective which was the positional cloning of the *shd* gene. These phenotypic aspects of the study were significant for guiding the cloning project and most importantly provide a biological and functional framework to the whole project.

5.2 Materials and Methods

5.2.1 Whole mount in situ hybridisation

The procedure followed for mRNA in situ hybridisation on all mount zebrafish embryos was already described fully in Chapter 4 section 4.2.1.8. The probes for the following genes were used: zebrafish *shd/ALK* synthesised in this work, *rose/ednrb1* (Parichy et al., 2000a), *nac/mitf* (Lister et al., 1999), *dopachrome tautomerase* (*dct*) (Kelsh et al., 2000b), *fkd6/foxd3* (Odenthal and Nusslein-Volhard, 1998).

5.2.2 Antibody staining of dorsal root ganglia and enteric ganglia

Antibody staining with anti-Hu, mAb 16A11 (Marusich et al., 1994) was performed on tissue cryosections, whole mount embryos and dissected guts.

5.2.2.1 Tissue cryosections

Agar embedding for cryostat sectioning of fixed embryos was done by melting an aliquot of agar-sucrose solution (1.5% agar, 5% sucrose) and cooling it to 40°C. Animals were transferred with a drop of fix buffer to a flexible plastic well mould. Excess liquid was removed and the agar-sucrose was added to cover the embryos. Using fine needles the embryos were oriented properly and the agar was set aside until it was solidified (5-10 minutes). Blocks were removed from the mould and trimmed with a razor blade according to the desired plane of sectioning. 2-3mm of agar were left around the tissue and 3-4mm below it to allow a surface where it would be frozen to the cryostat chuck. The trimmed agar blocks containing the animals were placed into a 30% sucrose solution and stored at

4°C until all blocks had sunk (few hours) indicating impregnation. A slightly raised platform was made by freezing a layer of OCT compound (Tissue Tek Miles) to the cryostat chuck. A drop of OCT was added to the platform and the agar blocks positioned and lowered slowly into a container of liquid nitrogen. Even freezing was necessary to avoid cracking the blocks. These were equilibrated to the cryostat temperature before sectioning. Sections 10-40μm thick were cut with this procedure and dried on a 26°C hot plate.

For the antibody staining procedure an Immedge pen (Vector) was used to form a hydrophobic barrier around the slide. Sections were rinsed briefly in PBS (tablets from Oxoid), and pre-incubated in PBDTX containing 2.5% of sheep serum for at least 30 minutes at room temperature. PBDTX contains: 1g BSA (Sigma),1ml DMSO (Sigma), 0.5ml 10% Triton X-100 (Sigma) and distilled water to make a total volume of 100ml. The pre-incubation solution was removed and the primary antibody (1:3000), mouse anti-Hu, was added in PBDTX containing 2.5% of sheep serum for 2 hours at room temperature or over night at 4°C. Sections were then rinsed 3 times in PBS for 10-15 minutes each wash and subsequently the secondary biotinylated antibody (1:200), goat anti-mouse (Roche), was added in PBDTX containing 2.5% of sheep serum for 1 hour at room temperature or overnight at 4°C. This was also rinsed 3 times in PBS for 10-15 minutes each wash.

Vectastain Elite ABC reagent (Vector) was diluted by adding 20μl reagent A per ml to PBDTX, followed by 20μl reagent B, mixed and let to stand for 30 minutes at room temperature. The diluted reagent was added to the sections for 1 hour at room temperature or overnight at 4°C after which it was washed 3 times in PBS for 10-15 minutes. 3,3'-Diaminobenzidine (DAB) substrate, from the DAB substrate kit for peroxidase (Vector), was diluted in PBSX (14μl/ml) and the sections incubated for 15 minutes. Then, 7μl hydrogen peroxide (Sigma) were added per 1ml diluted DAB solution. The reaction (usually very quick) was closely monitored under a dissecting scope and stain

development stopped by rinsing several times with PBS. Slides were coversliped in 50% glycerol: PBS solution and analysed under a Eclipse E800 microscope (Nikon).

5.2.2.2 Dissected guts and whole mount embryos

Whole mount embryos up to 60hpf or dissected guts of older embryos were stained using this slightly different protocol. Embryos to be antibody stained were dechorionated and fixed in 4% paraformaldehyde/PBS overnight at 4°C and then kept in PBSX (0.5% Triton-X (Sigma) in PBS) at 4°C until needed. They were washed 2x5 minutes with PBSX, 4x30 minutes in milliQ water, incubated in acetone for 7 minutes at –20°C and blocked for at least one hour in PBSDX (50ml 2xPBS pH7.3, 1g BSA (Sigma), 1ml DMSO (Sigma), 0.5ml 10% Triton X-100 and milliQ water up to 100ml) containing 15µl/ml horse serum (Sigma) block. Primary antibody, anti-Hu was diluted 1:5000 in PBSDX containing 15µl/ml horse serum and the embryos incubated overnight at 4°C. On the following day embryos were washed 4x30 minutes in PBSDX, the biotinylated secondary horse anti-mouse/rabbit antibody diluted 1:5000 in PBSDX with 15µl/ml horse serum and the embryos incubated overnight at 4°C.

The antibody was discarded, the embryos washed 4x30 minutes in PBSDX, incubated 45 minutes in the diluted ABC reagent (Vectastain ABC Elite kit), washed 3x30 minutes in PBSDX and 30 minutes in PBSX. The DAB substrate was diluted in PBSX (14µl/ml) and the embryos incubated for 15 minutes. After transferring the embryos to a glass well plate 7µl hydrogen peroxide (Sigma) were added per 1ml diluted DAB solution. The coloured precipitate was developed in the dark for 5 minutes up to a few hours. For long incubations it was necessary to add more hydrogen peroxide approximately every 10 to 15 minutes. The coloured precepitate development was finally stopped by rinsing the embryos several times in PBSX and embryos were kept at 4°C.

5.2.2.2.1 Using a fluorescent secondary antibody

This protocol was used for whole mount embryos older than 60hpf since the sensitivity of detection of structures deep within the tissue is greater with the fluorescent detection than with the DAB substrate.

Embryos were washed 3x5 minutes at room temperature in PBSX, 3x1 hour in milliQ water, pre-incubated in 0.75ml block [0.5% Triton X-100, 1% DMSO, 5% horse serum in PBSX] and incubated at room temperature overnight with the primary anti-Hu mouse antibody diluted 1:1000 in 0.75ml block. The antibody was discarded, embryos rinsed in PBSX as described in the previous protocol, washed for 3x1 hour in PBSX and incubated at room temperature overnight with the secondary antibody, Alexa Fluor 546 rabbit anti mouse IgG (Molecular probes, Oregon), diluted 1:800 in 0.75ml block. Again, the antibody was discarded, embryos rinsed in PBSX, washed for 3x30 minutes in PBSX, transferred into 50% glycerol/PBS and incubated for 15 minutes. Embryos were analysed with an Eclipse E800 microscope (Nikon) using fluorescence microscopy (TRITC filter).

5.2.3 Transplant experiments

WT into *shd* chimaeras were obtained as follows: rhodamine conjugated biotinylated dextran dye (Molecular Probes) was injected into the yolk of 2 cell stage wild-type embryos (0:45 hpf) using a Drumond Nanoinjector and a dissecting scope. Embryos were then allowed to develop until 30% epiboly stage (5:50 hpf) according to (Kimmel et al., 1995). At this stage wild-type cells were transferred to an unlabeled shield stage host (6 hpf). Each slide had one donor and three hosts mounted on cold methylcellulose. Hosts were derived from *shd*$^{+/-}$ x *shd*$^{+/-}$ crosses and thus were a mixture of WT, *shd*$^{+/-}$ and shd$^{-/-}$ embryos. Donor cells were placed into host embryos in the approximate position of the neural crest as judged from the fate map at the same developmental stage (Kimmel et

al., 1995). The total number of surviving embryos at 72 hpf was recorded. As expected, approximately 25% of hosts proved to be *shd* $^{-/-}$. These *shd* $^{-/-}$ embryos were examined under fluorescent light for the presence of rhodamine labelled cells and thus scored as chimaeras. Chimaeric *shd* $^{-/-}$ embryos we examined for the presence of reflecting iridophores under incident light. All rhodamine labelled iridophores were scored using a rhodamine filter block on an Elite E800 (Nikon) microscope. All other cell types in the chimaeric *shd* $^{-/-}$ embryos were also scored for rhodamine label.

In addition to the rhodamine labelling, confirmation of the donor cell origin was performed by detection of biotin using an anti-biotin secondary antibody following the protocol described below. This confirmation was needed because iridophores are auto fluorescent under the same channel as rhodamine and sometimes this distinction can be difficult.

The following detection procedure was applied to all chimaeric mutant embryos. Embryos were fixed in 4% paraformaldehyde/PBS overnight at 4°C or for 1-2 hours at room temperature. Washed 3x 5minutes with 0.1M phosphate buffer at room temperature. [Phosphate buffer: 80ml 0.1M Na_2HPO_4; 20ml 0.1M NaH_2PO_4, pH 7.3] and then 4x30 minutes with distilled water at room temperature. Embryos were cooled at –20°C for 7 minutes in cold acetone and susequently washed for 5 minutes with 0.1M phosphate buffer at room temperature. Background peroxidase activity in blood cells was quenched by incubating the embryos for 30 minutes in methanol with 5% hydrogen peroxide. Embryos were washed 3x 5 minutes with 1% PBDTX and incubated for 2 hours with Ready to Use (RTU) Vectastain ABC (Vector). They were then washed for 3x 15 minutes with 1% PBDTX and pre-incubate with DAB (14µl per ml PBS/0.1% Triton X-100) for 15 minutes at room temperature. 7µl H_2O_2 were added per ml DAB solution and mixed well. Stain development was stopped when desired by washing several times with PBSX and storing at 4°C for at least overnight.

5.3 Results

5.3.1 Iridophores in wild-type and *shd* $^{-/-}$ embryos

To understand the *shd* phenotype it is important to describe in some detail the wild-type early larval iridophore pattern. In wild-type embryos the iridophores are positioned among the melanophores making it difficult to photograph them. Hence, we have used *sandy* (*sdy*) homozygous embryos which lack the melanin pigment but are otherwise normal (Kelsh et al., 1996) to allow us to fully visualise the wild-type iridophore pattern both with incident and transmitted light.

The embryonic iridophore pattern is complete by 5 dpf (Kelsh et al, 1996). Iridophores are discernible in the chorion lying over the retinal epithelium melanophores as early as 3 dpf. In 3 dpf *sdy* embryos it is possible to see iridophores with transmitted light. They form a spotted pattern over the eyes, densest around the iris where they form a continuous ring (Figure 5.2a). Iridophores are also present in three of the four embryonic stripes. At 3 dpf they can be seen in the dorsal, ventral and yolk sac stripes. These are respectively, dorsal to the neural tube, dorsal to the gut and ventral to the yolk. In the dorsal stripe they aggregate in a large patch on the top of the head and as a row of discrete spots along the trunk to the tip of the tail (Figure 5.2c). These spots are usually made of a few cells together and they are spaced fairly evenly. Occasionally, by 3 dpf onwards they can be fused forming a thin continuos stripe of iridophores. In the ventral stripe iridophores form large patches, one on each side of the embryo, over the yolk sac, known as the lateral patches (Figure 5.2b). These become narrower towards the back of the embryo and are replaced by the characteristic double spotted stripe of iridophores that run until the tip of the tail where they meet the dorsal stripe. Iridophores can also be found as a flat sheet covering the region under the yolk sac and yolk sac elongation. Here they are not spotted

or patchy, in contrast, they appear as a thin reflective layer of cells in tight association with the yolk sac stripe melanophores. Iridophores are characteristically compact and round or oval in shape. However, some migratory differentiated iridophores are seen (e.g. over the yolk patches melanophores) and these have the typical elongated or branched shape of a migratory neural crest cell. Under transmitted light or with DIC optics they appear reminiscent of unpolished gold stones (Figure 5.2d). When incident light is used iridophores look like shiny iridescent patches or spots and their shape and texture is not always detectable because its hindered by their great reflectivity.

In a *shd* $^{-/-}$ embryo the number of iridophores varies from very reduced to completely absent depending on the strength of the allele being analysed. Consequently, the patterns we have just described are very compromised. The patches or spots seen in wild-types, if not absent, are usually smaller (have less individual cells). However, the cells that are present appear normal in shape and brightness. In this project we have looked at three of the *shd* alleles: *ty82*, *ty9* and, *terminally dull* (*trd*) identified in a later screen (David Raibles' laboratory, Seattle). *ty82* is the strongest allele and only very rarely are one or two iridophores seen in embryos homozygous for this mutation (Figure 5.2 k,m). These are usually on the interior side of the eye or in a yolk patch. This mutation is lethal because the mutants never develop a swim bladder. In contrast, the *ty9* homozygotes are viable. Their embryonic iridophore pattern is compromised but they often have smaller patches of cells in the eyes (Figure 5.2 e,f,g) or on the lateral patches and can have incomplete spot patterns in the dorsal and ventral stripes. These mutants develop a swim bladder and, most interestingly around the transition from larvae to adult stage (4 weeks after fertilisation) seem to recover a normal iridophore pattern (figure 5.2h) and develop into normal looking adults. The *trd* homozygous mutants, as hinted by their name, also die without forming a swim bladder. These mutants have a phenotype that lies somewhere in between *ty9* and *ty82*, perhaps more similar to the *ty82* phenotype (Figure 5.2 i).

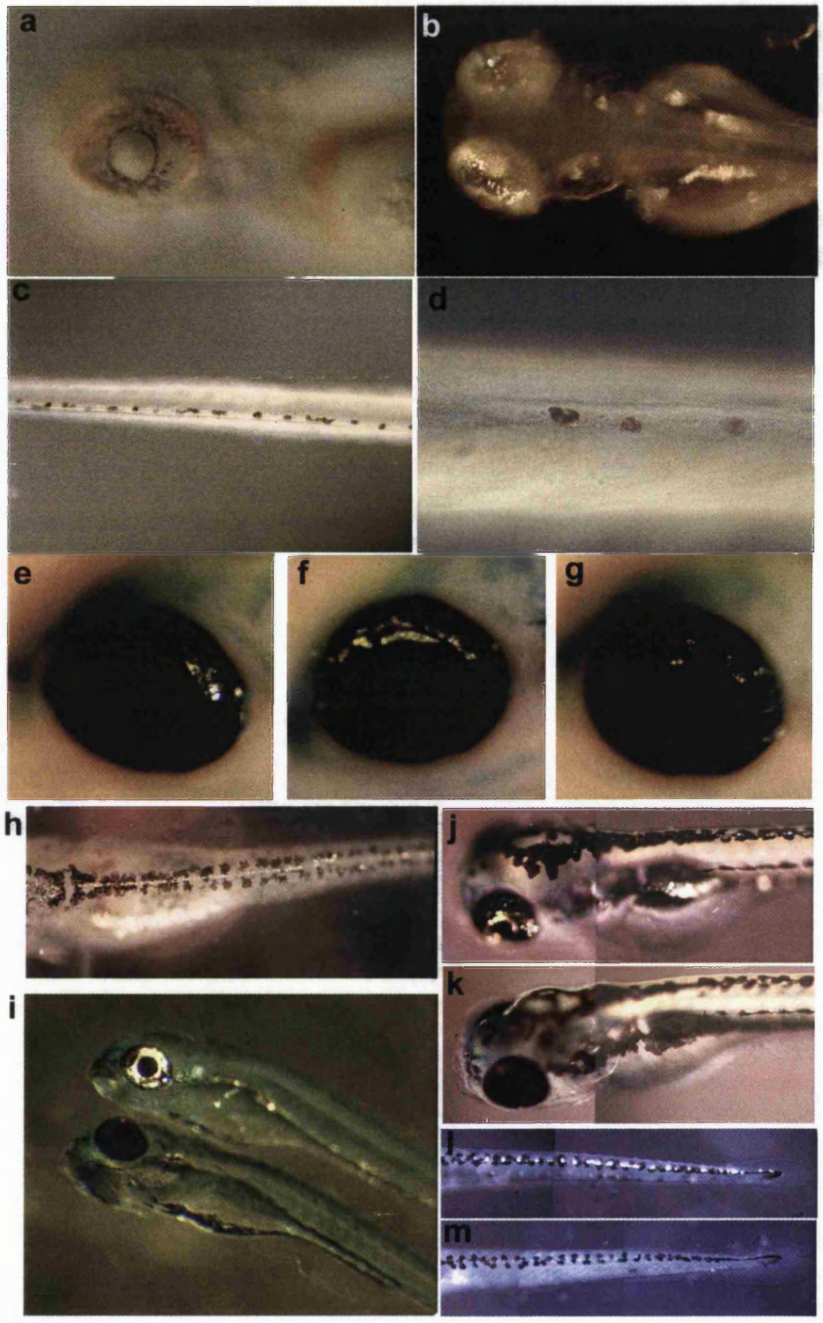


Figure 5.2 Embryonic iridophore phenotype. (a-d) sdy homozygous embryos at 72 hpf. (a) Iridophores can be seen in the retina around the iris with transmitted light. (b) dorsal view of the head focusing on reflective iridophores in the eyes and lateral paches above the yolk sac. (c) iridophores on the dorsal stripe under transmitted light. (d) higher magnification of the dorsal stripe showing the oval/round shape and texture of individual cells.(e,f,g) *ty9* homozygous embryos at 72 hpf. Different strength eye phenotypes that can be found in these mutants. (h) *ty9* homozygous mutant embryo 20 dpf at transition stage from larvae to adult showing recovered iridophore pattern.(i) *trd* homozygous mutant embryo (bottom) and wild-type sibling (top) at 72 hpf. (k) *ty82* homozygous mutant embryo showing complete lack of iridophores in the eye and lateral pach compared to its wild-type sibling (j). (m) *ty82* homozygous mutant embryo without iridophores in the dorsal stripe compared to its wild-type sibling at 72 hpf (l), orientation is as in other Figures.

5.3.2 Marker genes for the Neural Crest pigment lineages

5.3.2.1 Identification of *ednrb1* as a marker for the iridophore lineage

rose mutants have wild-type early larval pigment patterns but fail to develop normal numbers of melanophores and lack iridophores in the adult (Parichy et al., 2000a). *ednrb1* is the gene responsible for the *rose* mutant phenotype and encodes a G-protein-coupled receptor with seven transmembrane domains (Parichy et al., 2000a) (see Chapter 1 for Figure of this phenotype). In order to test whether *ednrb1* was expressed in embryonic melanophores and iridophores in spite of the lack of a *rose* embryonic phenotype, we performed a time course in situ hybridisation. We identified the *ednrb1* gene as a marker for iridoblasts (48 hpf) and iridophores (72hpf). However, before 48hpf this gene seems to be expressed in the other pigment lineages as well and therefore it is not suitable as a early iridophore lineage marker (Parichy et al., 2000a).

Using mRNA in situ hybridisation we examined expression of *ednrb1* in *shd* $^{-/-}$ and wild-type embryos between 12 and 72 hpf (Figure 5.3). Lack of *ednrb1* expression in *shd* $^{-/-}$ was very dramatic at 72 hpf. At this stage iridophores are readily visible under incident light and their absence in *shd* $^{-/-}$ was obvious (Figure 5.3 g,h). At this same time-point the expression of *ednrb1* followed the same pattern as iridophores. In wild-type embryos labelled cells were detected in the correct locations and patterns for iridophores; these were completely absent from *shd* $^{-/-}$ embryos (Figure 5.3 i,j). Labelled presumed iridophores could thus clearly be seen in a 72 hpf *sdy* $^{-/-}$ embryo without the melanin pigment obscuring pattern (Figure 5.3 k) and in transversal histological sections of wild-type embryos (Figure 5.3 l). The reduced expression of *ednrb1* in *shd* $^{-/-}$ embryos could be traced earlier to 24 hpf in a region dorsal to the eye when compared to wild-type siblings (Figure 5.3 c,d). At 30 hpf there were distinct expression patterns around the otic vesicle

and again *shd* $^{-/-}$ embryos had less labelled cells than their wild-type siblings (Figure 5.3 e,f) . At these early stages mutants were identified on the basis of a difference in phenotype seen in 25% of the embryos resulting from a heterozygous cross. No differences were found between mutants and wild-types at 12 and 18 hpf (Figure 5.3 a,b).

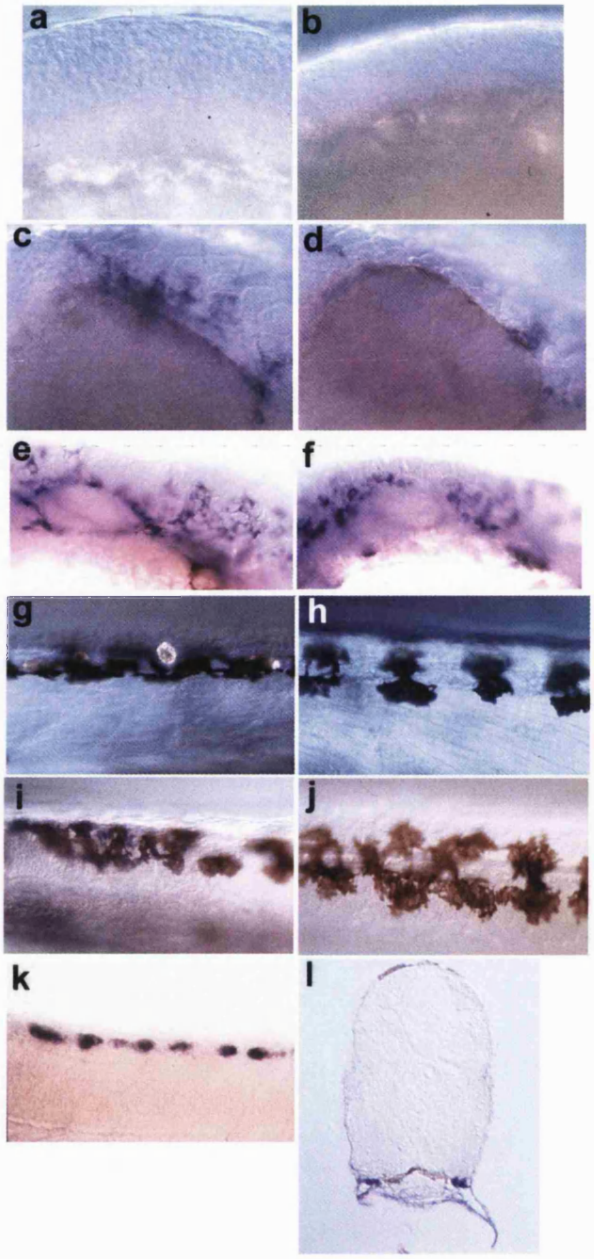


Figure 5.3 *ednrb1* expression in embryonic and early larval development. *ednrb1* is not expressed in 12 nor 18 hpf zebrafish embryos, respectively (a,b). At 24 hpf *shd* $^{-/-}$ show lower levels of expression of *ednrb1* around the eye region (d) when compared to wild-type siblings (c). The same is seen at 30 hpf around the otic vesicle of *shd* $^{-/-}$ (f) and wild-type (e). At 72 hpf the iridophores are readily visible among the melanophores of the dorsal stripe using incident light (g). *shd* $^{-/-}$ do not have iridophores but have normal melanophore pattern (h). *ednrb1* expression localises to cells (purple) that are in the right position to be iridophores (i) these are absent in *shd* $^{-/-}$ (j). Presumed iridophores labelled with *ednrb1* in a 72 hpf *sdy* $^{-/-}$ without melanin pigment for better visualisation of the iridophore pattern in the dorsal stripe (k). Histological transverse section showing ventral stripe iridophore spots on each side of the embryo (purple) (l).

5.3.2.2 Are the melanophore and xanthophore lineages affected in *shd*$^{-/-}$ embryos?

shd $^{-/-}$ embryos do not show any abnormalities in melanophore or xanthophore pattern or shape. In order to confirm that these other pigment lineages were unaffected in *shd* $^{-/-}$ embryos we used two molecular markers, one for the melanophore lineage, *dopachrome tautomerase* (*dct*) (Kelsh et al., 2000b) and an other for the xanthophore lineage, *GTP-cyclohydrolase* I (*gch*) (Parichy et al., 2000b). Melanophores are labelled naturally by their melanin pigment from 27 hpf. *dct* is a marker for melanoblasts and melanophores because it was shown to label undifferentiated cells (5 hours before melanisation) as well as differentiated cells (melanised) (Kelsh et al., 2000b) . Comparison of *dct* expression pattern in *shd*$^{-/-}$ and wild-type siblings shows that there are no detectable differences (data not shown). It was therefore concluded that both melanoblasts and melanophores seem to be unaffected in *shd* $^{-/-}$ in terms of number, pattern and shape.

The xanthophore lineage marker *gch* converts guanosine triphosphate to tetrahydrobiopterin (BH4), a precursor to pteridine pigments. BH4 also is a co-factor for phenylalanine hydroxylase, which converts phenylalanine to tyrosine (a precursor to melanin), and regulates the activity of tyrosinase, which converts tyrosine to melanin intermediate, dopa. Thus, *gch* should be expressed by xanthoblasts and possibly melanoblasts (Parichy et al., 2000b). The time points chosen for the in situ hybridisation with this probe were 30, 49, 54 and 77 hpf (on PTU treated embryos for better visualisation). Again the expression pattern of this marker on *shd* $^{-/-}$ showed no distinction from wild-type siblings.

In order to test if *gch* was expressed by iridoblasts we used the zebrafish *shd/ALK* probe (see Chapter 4) in parallel to compare the expression patterns at 30 and 49 hpf. Although, double in situ hybridisations were not performed it was clear from the very distinct expression patterns that xanthoblasts and iridoblasts do not share expression of these

markers. The dorsal stripe pattern of iridoblasts, described earlier as spots of round/oval cells positioned on top and along a single dorsal line, is not comparable to the extremely branched pattern of *gch*+ cells. Furthermore, *gch* expressing cells are positioned surrounding but not intermixed with the positions where iridoblasts are seen (Figure 5.4 a). The dorsal branched pattern extends ventrally throughout the trunk and tail and is not confined to stripes as iridoblasts do (Figure 5.4 b). Again the expression of *gch* in the eye (Figure 5.4 c) is very different from the *shd/ALK* expression at 48-49 hpf. *gch*+ cells are branched in shape not round, they are often very large with thin projections, never form a spotted pattern, and do not surround the iris as iridoblasts do (for the *shd/ALK* expression pattern see Chapter 4).

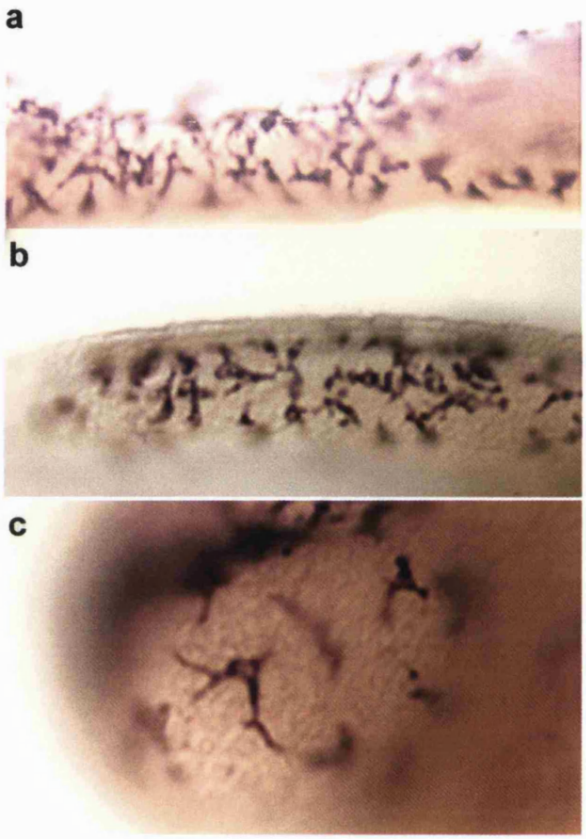


Figure 5.4 *gch* expression pattern in *shd* $^{+/-}$ and *shd* $^{-/-}$ embryos is the same. (a) dorsal view of the trunk of a 30 hpf showing the extremely branched pattern of *gch+* cells. (b) trunk lateral view of a 30 hpf embryo showing that *gch+* cells do not assemble in any dotted-stripe pattern in contrast they extend their projections all over the body. (c) eye of a 49 hpf embryo showing branched *gch+* cells all over the retinal epithelium. Embryos were treated with PTU to avoid melanin visualisation.

5.3.2.4 *foxD3* as a marker for iridoblasts

FoxD3 belongs to the winged-helix or forkhead class of transcription factors and has been identified as a marker for pre-migratory neural crest, floorplate, somites and tailbud (Odenthal and Nuesslein-Volhard, 1998). More recently Kelsh et al. (2000) identified it in neural crest-derived glia in the cranial ganglia and along the posterior lateral line nerve. More interestingly, J. Lister (personal communication) identified expression of *foxD3* in iridophores and has evidence that this transcription factor might be involved early in iridophore lineage specification. To test this hypothesis we decided to perform some preliminary *foxD3* expression studies on *shd* $^{-/-}$ to test whether this gene (1) might have a role in the iridophore lineage and (2) in particular, at what stage of development *foxD3* was acting: upstream or downstream of *shd/ALK*?

Results from preliminary expression studies done at 30 and 49 hpf showed no differences between *shd* mutants and wild-type siblings. However, a subset of *foxD3*+ cells were positioned in the right locations to be identified as potential iridoblasts both at 30 and 49 hpf. At 49 hpf we observe strong *foxD3* expression in the cranial ganglia associated glia, lateral line glia (Figure 5.5a) and in cells identified as iridoblasts because of their position and spacing on the dorsal and ventral stripes (arrows in Figure 5.5b). This expression pattern strongly resembles the one obtained with *shd/ALK* riboprobe (see Chapter 4) and suggests that *foxD3* may indeed play a role in the iridophore lineage.

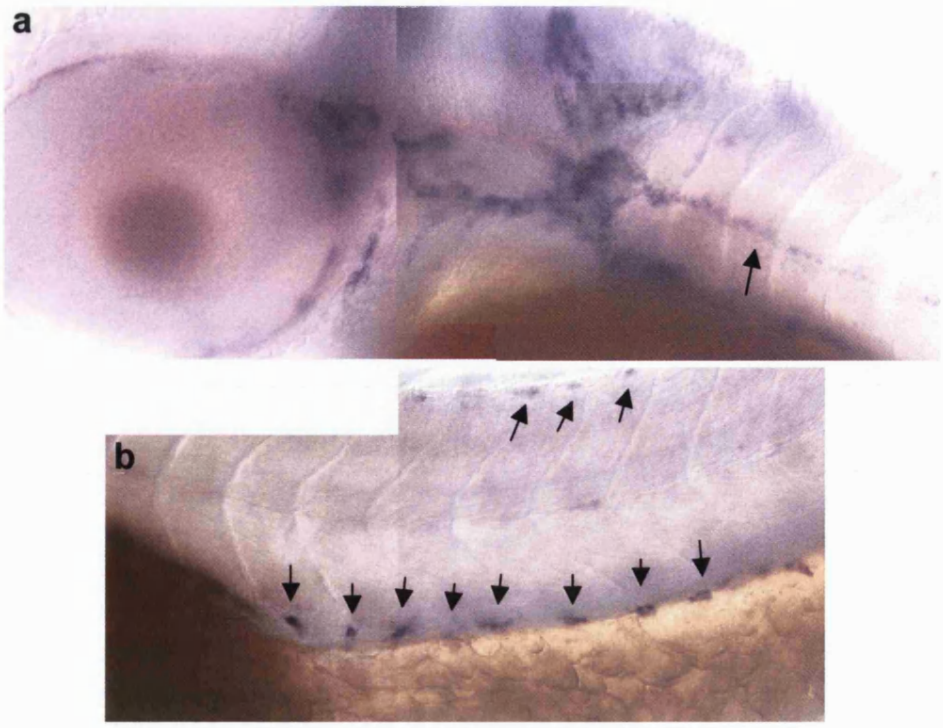


Figure 5.5 *foxD3* has an intriguing expression pattern in all embryos from an heterozygous *shd* cross at 49 hpf. (a) *foxD3+* cells in cranial ganglia and lateral line associated glia are still present at 49 hpf. (b) lateral view of the trunk showing *foxD3+* cells on the dorsal and ventral stripes (arrows) positioned in a pattern reminiscent of that of iridoblasts when labelled with *shd/ALK*.

At 30 hpf *foxD3* expression is seen in the glia associated with the cranial ganglia (Figure 5.6a arrowheads), in the glia associated with the lateral line nerve (thin arrows) and also in some cells that are in the position of iridoblasts above the yolk sac elongation (single arrow). On the medial pathway (Figure 5.6b) *foxD3+* cells are seen throughout the trunk but only a subset of these are closer to potential iridoblast positions above the yolk (arrows in Figure 5.6b) and on the dorsal stripe. Some of these cells could also be melanophores based on their branched shape but more likely, these cells can be medially migrating neural crest cells that are not yet fate restricted. Other cells, on a more superficial focal plane, positioned above the yolk sac and yolk sac elongation are potential iridoblasts (Figure 5.6c) that seem to be migrating along the ventral stripe.

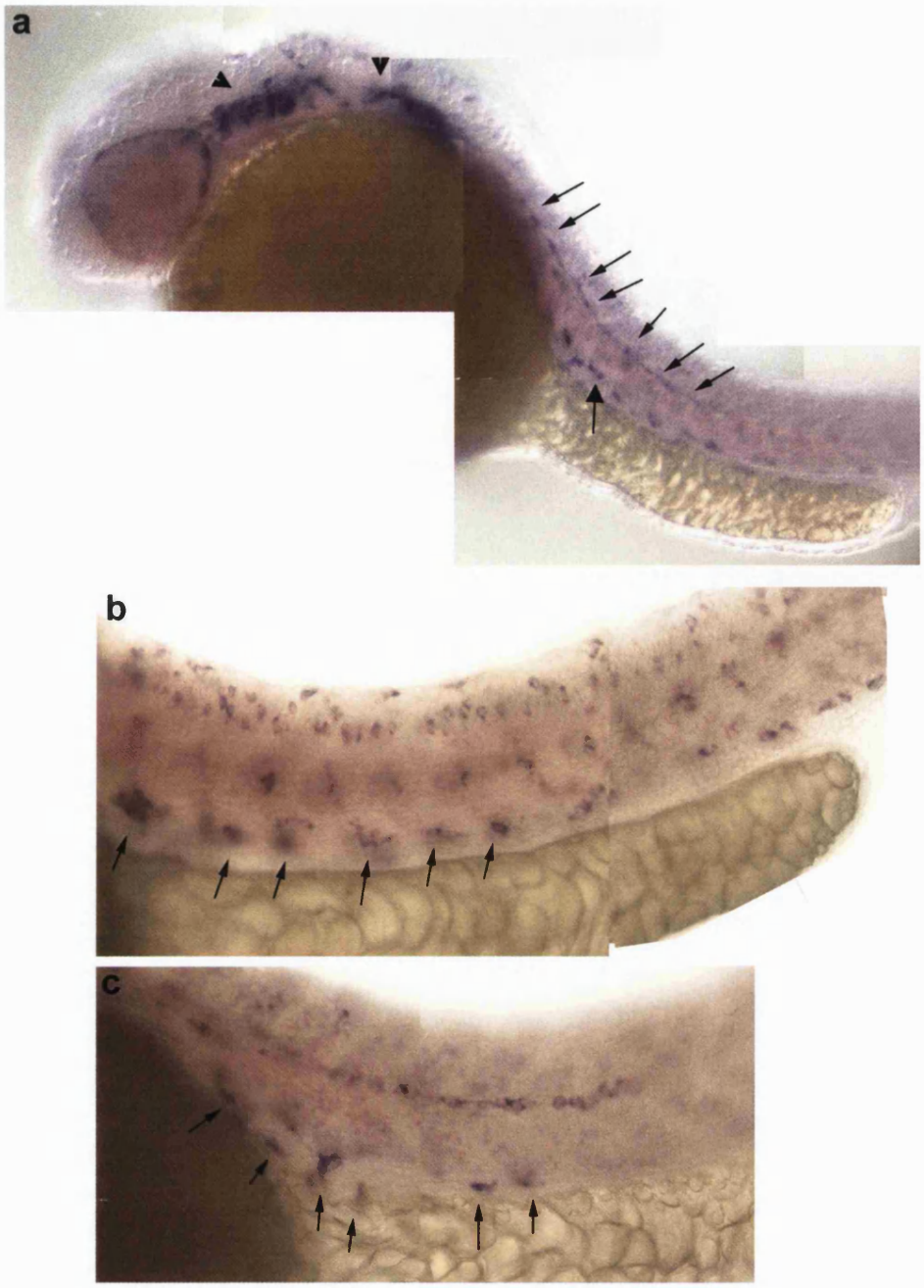


Figure 5.6 *foxD3* expression in all embryos resulting from a heterozygous *shd* cross at 30 hpf. (a) *foxD3* expression is seen in the region of the cranial ganglia (arrowheads), lateral line nerve (thin arrows) and in potential iridoblasts (broad arrow) on a superficial focal plane. (b) *foxD3+* cells on the medial pathway are spread throughout the trunk but only a subset of these are closer to iridoblast positions (arrows). (c) Superficial focal plane showing *foxD3+* cells on lateral line nerve and in potential iridoblasts on the ventral stripe above the yolk sac (arrows).

5.3.3 Genes that might interact with *shd/ALK*

5.3.3.1 *nacre/mitfa*

nacre homozygous mutants (*nac*$^{-/-}$) are characterised by their lack of melanophores and their precursors throughout development. In contrast, these mutants were reported to have an increase in iridophore number (Lister et al., 1999). The *nacre* gene was identified as encoding a basic helix-loop-helix/leucine zipper transcription factor related to *microphthalmia* (*Mitf*), a gene known to be required for development of eye and neural crest-derived pigment cells in the mouse. In order to investigate iridophore development in *nac*$^{-/-}$ we used the iridoblast and iridophore markers *ednrb1* and *shd/ALK* on *nac*$^{-/-}$ embryos. We asked whether *nac*$^{-/-}$ embryos have more iridoblast precursors.

Surprisingly, both *ednrb1* and *shd/ALK riboprobes* showed delayed expression in iridoblasts in *nac*$^{-/-}$ embryos. Thus, at 48 hpf *nac*$^{-/-}$ embryos did not express *ednrb1* (Figure 5.7b) whilst, as noted before, their wild-type siblings already did (Figure 5.7a). At 56 hpf the expression in *nac*$^{-/-}$ embryos was still very faint compared to their wild-type siblings (Figure 5.7c,d). Only at 72 hpf were the levels of *ednrb1* expression considered equivalent between *nac* $^{-/-}$ and siblings (Figure 5.7e,f). It is worth stressing that these embryos were processed and developed in parallel in the same tube throughout the protocol. Similar observations were made with the expression of *shd/ALK* at 30 hpf. While wild-type embryos showed iridoblast probe expression in both dorsal (Figure 5.7g) and ventral (Figure 5.7h) stripes, expression was completely absent in *nac*$^{-/-}$ embryos (Figure 5.6i). However, by 48 hpf *shd/ALK* expression in iridoblasts was identical in both *nac* $^{-/-}$ and *nac*$^{+/-}$ embryos (data not shown). In summary, *ednrb1* expression, usually detectable very clearly at 48 hpf in *nac*$^{-/-}$ is still faint at 56 hpf and *shd/ALK*, usually very strong at 30 hpf was only detected at 48 hpf in *nac*$^{-/-}$ (but is possible that it could be visible sometime earlier between 30 and 48 hpf).

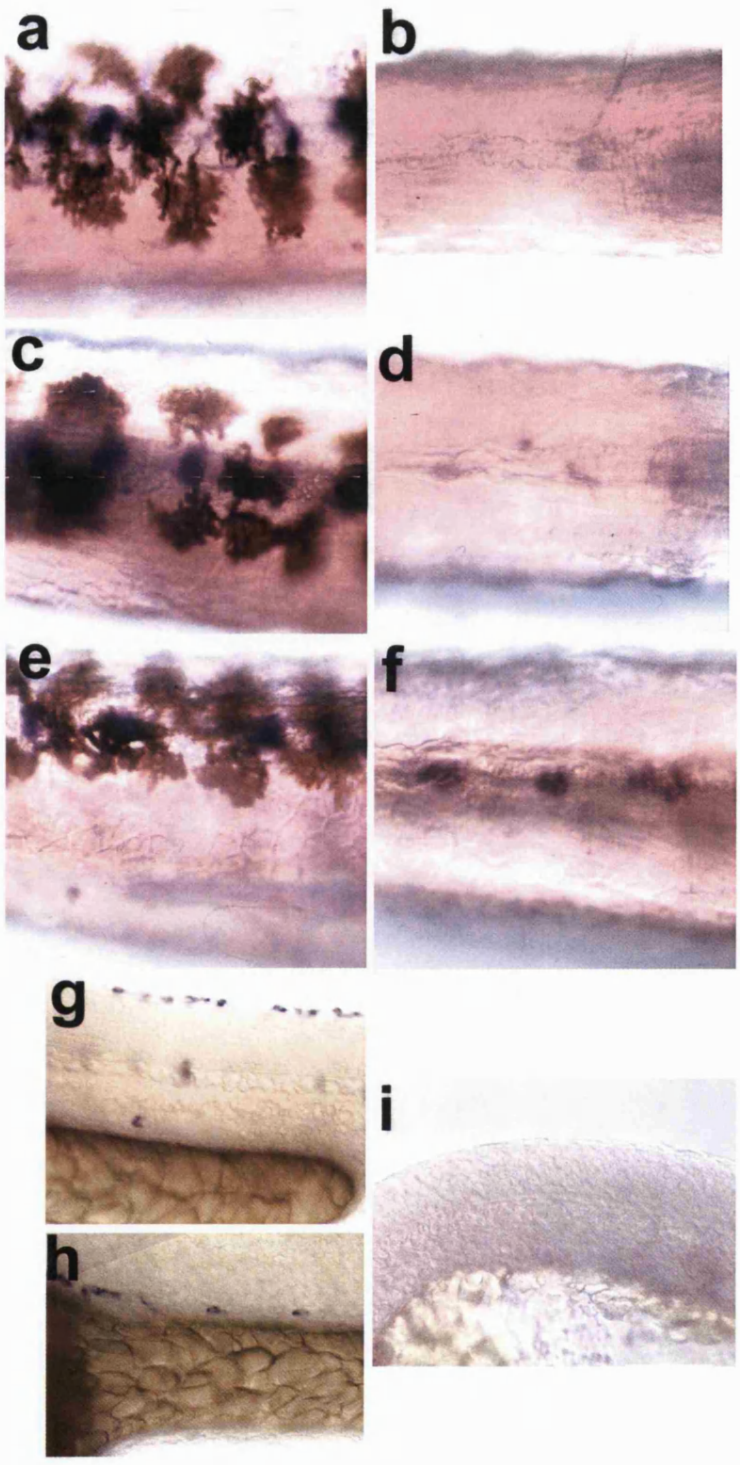


Figure 5.7. Expression of iridoblast markers in *nac*$^{-/-}$ embryos is delayed. (a) *ednrb1* expression at 48 hpf in wild-type and (b) its absence in *nac* $^{-/-}$ siblings. (c) *ednrb1* expression at 56 hpf in wild-type and (d) very faint expression in *nac* $^{-/-}$ siblings. (f) Normal levels of *ednrb1* expression in *nac* $^{-/-}$ embryos were only detected at 72 hpf compared to their wild-type siblings (e). *shd/ALK* expression in iridoblasts on the dorsal (g) and ventral (h) stripe of a wild-type PTU treated 30 hpf embryo compared to a *nac* $^{-/-}$ embryo of the same age (i) (note that (h) is more anterior to (g) and (i) is comparable to (h).

To further investigate the relationship between *shd* and *nac* function we created *nac*$^{-/-}$/*shd*$^{-/-}$ double mutants. The resulting phenotype was additive (data not shown). The double mutants lacked all melanophores as well as all iridophores at 72 hpf. This genetic result suggests that these mutations affect distinct and independent populations of pigment cells. Thus, *nac*$^{-/-}$ iridophores require *shd/ALK* function to form and *shd* $^{-/-}$ melanophores need the *nac/mitfa* function to develop.

5.3.3.2 *colourless/sox10*

cls homozygous mutants have severe defects in most crest-derived cell types, including all three types of pigment cells, neurons and glia from the PNS. It has been shown that all three types of pigment cell are affected by *cls* mutations cell-autonomously (Kelsh and Eisen, 2000). The *cls* gene encodes a zebrafish *sox10* transcription factor and has been suggested to have a primary role in specification of non-ectomesenchymal crest derivatives (Dutton et al., 2001). In this way, it is proposed that *cls* is upstream of genes with critical roles in specific fates. Such a role has been demonstrated for melanophores (Elworthy et al., 2003) where *nac/mitfa* is activated by *cls/sox10*. It is therefore conceivable that such a similar relationship (perhaps not so direct) might occur between *cls/sox10* and *shd/ALK* in iridophores. To test this idea we examined *shd/ALK* expression in *cls*$^{-/-}$ embryos to test the prediction that *shd/ALK* expression would be absent in *cls*$^{-/-}$ embryos.

Results confirmed the predictions made: At 27 hpf *shd/ALK* expression on the cls $^{-/-}$ eye (Figure 5.8b) was completely absent, although iridophore expression was prominent in wild-type siblings (Figure 5.8a). At this stage we noted the absence of expressing cells above the yolk in the ventral stripe of *cls*$^{-/-}$ embryos (Figure 5.8d) in contrast to the presence of labelled cells in the wild-type siblings (Figure 5.8c). Wild-type expressing cells appeared to be migrating down through the medial pathway (Figure 5.8ef). Again this pattern was absent in *cls* $^{-/-}$ embryos (data not shown).

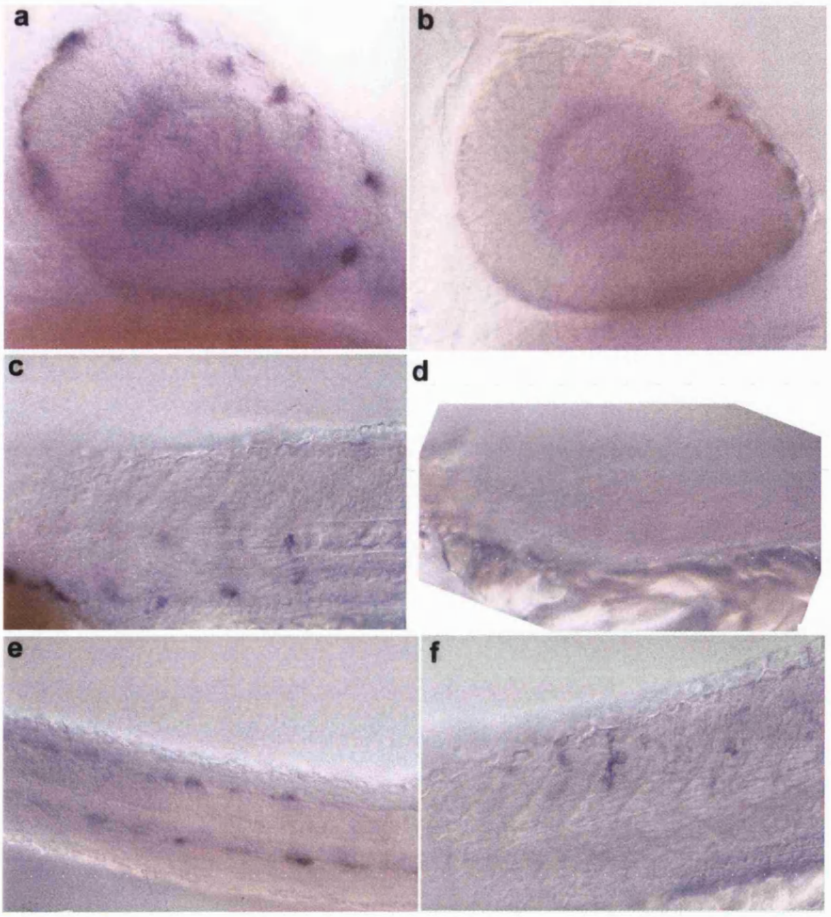


Figure 5.8 *cls*$^{-/-}$ embryos lack *shd*/*ALK* expression in neural crest derivatives. (a) 27 hpf wild-type eye showing expressing iridoblasts (b) absence of expressing cells in a *cls*$^{-/-}$ eye from the same cross. The pigmented cells in the *cls*$^{-/-}$ mutant eye are melanophores from the pigmented retinal epithelium that is not crest-derived. (c) Expressing iridoblasts in the ventral stripe of wild-type embryos in contrast to (d) absence of any labelled cells in a *cls*$^{-/-}$ sibling. (e-f) Iridoblasts at 27 hpf appear to be migrating through the medial pathway in wild-type embryos (e) dorsal view (f) lateral view.

5.3.3 Other neural crest derivatives in *shd*$^{-/-}$ are unaffected

We have shown that *shd*$^{-/-}$ embryos have a profound defect in the iridophore lineage but not on other pigment lineages. To ask whether other neural crest derivatives might be affected was particularly interesting because severe *shd* $^{-/-}$ have an unexplained lethality. We thus asked whether the enteric ganglia and dorsal root ganglia (DRGs) were normal by

using an anti-Hu immunostaining approach (Marusich et al., 1994). This antibody labels the cell bodies of differentiating neurons.

Our results showed no significant differences between mutants and wild-type embryos in respect to number, position and shape of both DRG neurons and enteric ganglia (Figure 5.9 a,b). We confirmed the observations for enteric ganglia made in whole mount embryos (Figure 5.9 c,d), in both dissected guts (Figure 5.9 e,f) and gut cryosections (Figure 5.9 g,h) using the peroxidase –antiperoxidase system. It is therefore concluded that the DRGs and enteric ganglia are normal in *shd* homozygous mutants and are not associated with the cause of lethality.

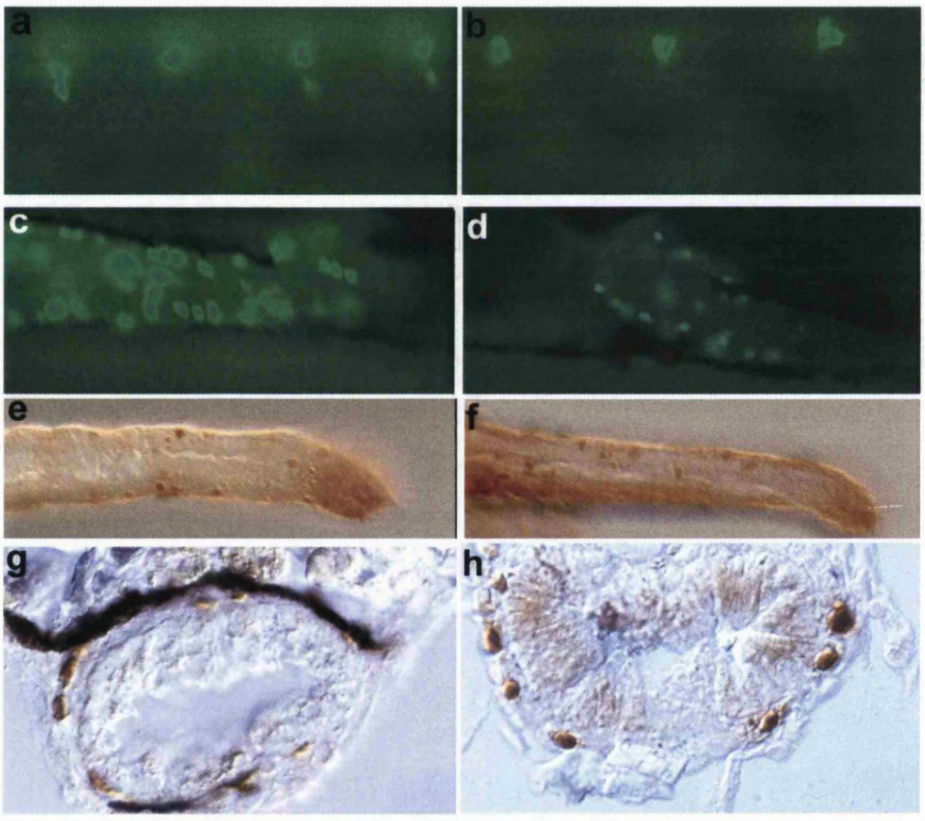


Figure 5.9 Neural crest derived dorsal root ganglia (DRGs) and enteric ganglia are normal in *shd* $^{-/-}$ embryos.(a,c,e,g) wild-type embryos at 5 days post fertilisation showing (a) DRG pattern and (c) enteric neuron pattern in whole mount immunostaning with anti-Hu. (b,d) Sibling mutant embryos show no detectable differences in these patterns, respectively. The same applies to dissected guts (e,f) and gut cryosections (g,h).

5.3.4 *shd* functions cell autonomously in iridophores

To learn whether *shd* mutations affect iridophores autonomously, cells were transplanted between wild-type and *shd* $^{-/-}$ embryos. Results from the mosaic analysis are illustrated in Figure 5.10 and presented fully in Table 5.1. We achieved good survival rates in our transplantation experiments with 64.6% survival of host embryos until 72 hpf when the *shd* $^{ty82/ty82}$ phenotype was first detectable. As expected, approximately, 25% of the embryos were *shd* $^{-/-}$. Among the 40 identified *shd* $^{-/-}$ embryos, about half (22) were chimaeric but only 7 of these had rescued iridophores. Iridophores were usually donor-derived; low level of occasional unlabelled iridophores reflects the normal *ty82* $^{-/-}$ phenotype. Rescued *shd* $^{-/-}$ were distinguished by exceptionally higher number of iridophores. When labelled iridophores were detected in an embryo these were at least 4. In some cases more than 20 were seen in the eyes, perhaps reflecting a high degree of proliferation in the eyes. The very few non-labelled iridophores detected in *shd* $^{-/-}$ chimaeric embryos do not seem to correlate with the position of the rescued ones suggesting that they were not involved in short distance signalling that could have enabled the wild-type iridophores to develop. Labelled cells in rescued chimaeras were sometimes restricted to iridophores (see experiment 3 on Table 5.1). Together, these data strongly argue for cell autonomy of *shd* function in iridophore development.

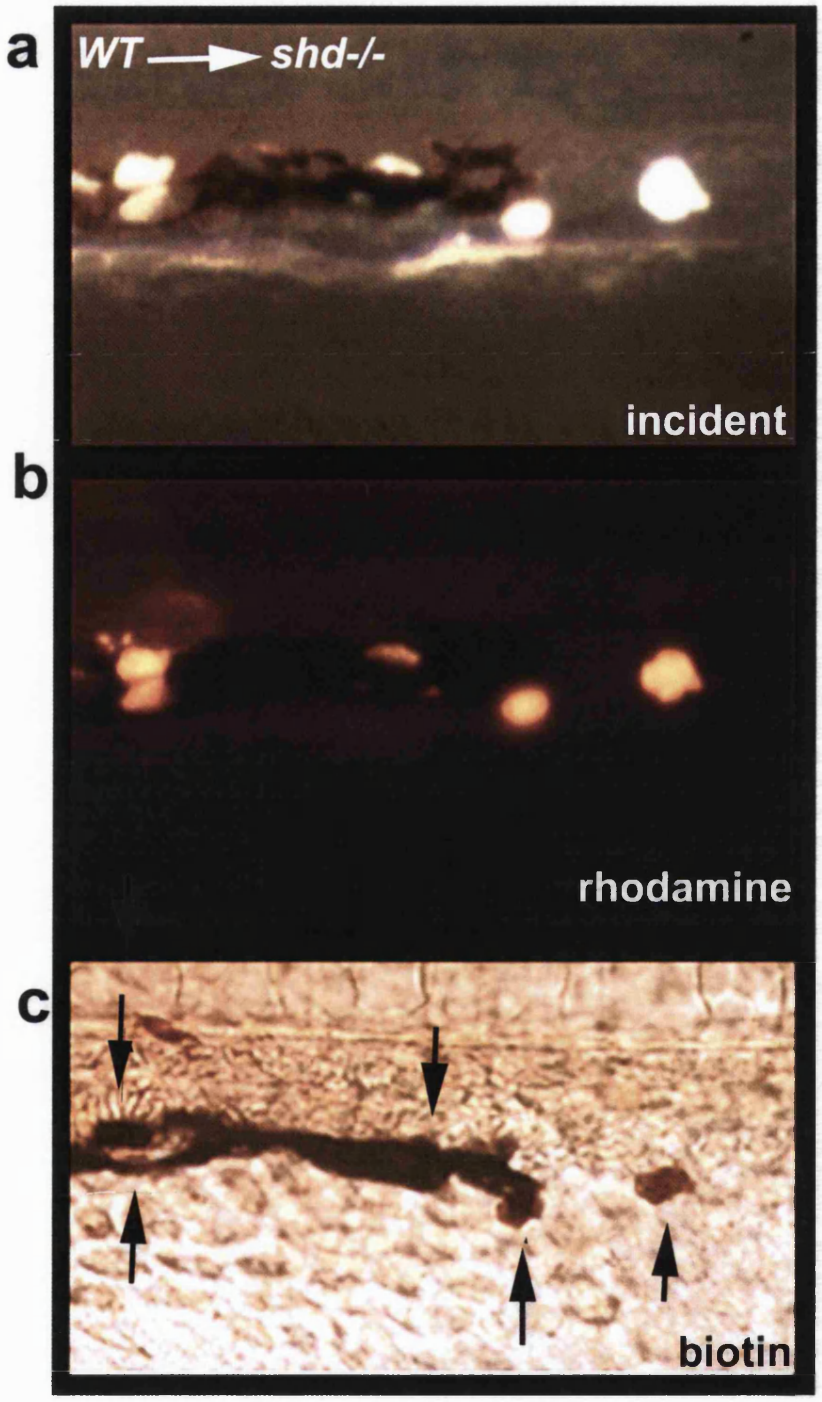


Figure 5.10 Rescued iridophores in wild-type → *shd* $^{ty82 / ty82}$ chimaeras are all donor-derived. The region shown is the same in the three panels and corresponds to a close up of the region of the ventral stripe near the tail of the chimaeric embryo described in experiment no. 8 (Table 5.1). 5 rescued iridophores are seen with incident light (a), and all show rhodamine fluorescence (b) and are thus donor (WT) derived. Confirmation of this is shown in panel c) where biotin-detection labels all 5 iridophores (in dark orange).

Table 5.1. Transplantation of labelled wild-type cells into *ty82* $^{-/-}$ embryos suggest *shd* functions cell autonomously in iridophores.

Exp. No.	No. hosts injected with WT cells	No. hosts survived until 72 hpf (%)	No. hosts identified as *ty82*$^{-}$ (%)	No. chimaeric *ty82*$^{-}$ embryos (labelled cell types)	No. chimaeric embryos with labelled iridophores	No. labelled iridophores (location in chimaeric embryo)	No. of non-labelled iridophores (location in chimaeric embryo)
1	19	19 (100)	5 (26.3)	5 (CNS; muscle)	0	0	0
2	16	13 (81.2)	3 (23)	1 (CNS; muscle)	0	0	0
3	38	28 (73.6)	5 (17.8)	2 (ventral and dorsal stripe iridophores; yolk)	2	5 (dorsal/ventral stripes) 4 (dorsal/ventral stripes)	1 (lateral patch)
4	20	12 (60)	3 (25)	0	0	0	0
5	34	9 (26.4)	2 (22.2)	2 (CNS; blood)	0	0	0
6	35	20 (57.1)	4 (20)	4 (CNS; melanophores; eye iridophores)	1	>20 (one eye)	0

Exp. No.	No. hosts injected with WT cells	No. hosts survived until 72 hpf (%)	No. hosts identified as *ty82*$^{-}$ (%)	No. chimaeric *ty82*$^{-}$ embryos (labelled cell types)	No. chimaeric embryos with labelled iridophores	No. labelled iridophores (location in the chimaeric embryo)	No. of non-labelled iridophores in chimaeric embryos
7	20	8 (40)	1 (12.5)	1 (CNS; muscle; skin; eye iridophores; lateral patch iridophores)	1	>20 (one eye) >20 (lateral patch on the same side as eye)	1 (dorsal stripe)
8	31	16 (51.6)	4 (25)	2 (CNS; ventral stripe iridophores; fin skin; PNS neuron)	1	9 (ventral stripe near the tail)	0
9	40	30 (75)	6 (20)	5 (melanophores; skin; stomach wall; eye and dorsal stripe iridophores; eye blood vessels; blood)	2	4 (dorsal stripe) >20 (one eye)	2 (lateral patch) 1(lateral patch)
10	38	33 (86.8)	7 (21.2)	0	0	0	0
total	291	188 (64.6)	40 (21.2)	22	7	~100	5

5.4 Discussion

From the in situ hybridisation and immunocytochemistry analyses presented here we propose that *shd* homozygous mutant phenotype seems to be specific to the iridophore lineage at least within the neural crest derived cell types. There is a strong decrease in the number of iridophores that varies with the allele being considered, but no effect was detected in other pigment cell lineages. Other neural crest derivatives such as dorsal root ganglia and enteric ganglia were also normal in *shd*$^{-/-}$embryos. We showed that *shd* functions cell autonomously in iridophores which was useful when it came to look at the expression pattern of *shd/ALK* (Chapter 4). We knew, from this result that a *shd* candidate gene should be expressed in the actual iridophore and/or its precursors. Taken together, these results confirmed that *shd/ALK* function in the developing neural crest is likely to be specific to the iridophore lineage. Furthermore, the almost complete absence of expression when using a *shd/ALK* probe in *shd*$^{-/-}$ confirms that *shd/ALK* functions very early in iridophore development. Thus, as it has been suggested earlier by Kelsh et al. (1996), *shd* represents a good candidate gene for the specification of iridophores. However, the preliminary observations of *FoxD3* transcripts in cells that occupy iridoblast positions even in *shd*$^{-/-}$ embryos make the specification argument less plausible. It is discussed here whether *shd/ALK* might still have a specification, a survival or an early differentiation role in the iridophore lineage from the neural crest.

The work already done on fish melanophores is important for helping us to distinguish the role of genes involved in specification vs. differentiation and survival. Dutton et al. (2001) suggest that *cls/sox10* has a primary role in specification of non-ectomesenchymal crest derivatives (pigment cells, neurons and glia from the PNS). Wnt signalling was shown to promote the formation of pigment cells at the expense of neurons and glia (Dorsky et al.,

1998) and hence specifies non-ectomesenchymal cells to a pigment cell fate. Other genes, downstream of *cls* and Wnt signalling would then be expected to specify individual lineages such as melanophores, xanthophores and iridophores (see Figure 5.11). This illustrates the principal of progressive fate restriction reviewed by (Kelsh and Raible, 2002). Of these three pigment lineages the best studied is that of melanophores. For instance we know that *cls/sox10* directly activates *nac/mitfa* (Elworthy et al., 2003) probably together with the previous demonstrated activation by the Wnt signalling (Dorsky et al., 2000) and that *nac/mitfa* is necessary and sufficient to make melanophores (Lister et al., 1999). Therefore, we assign the melanophore specification event to *nac/mitfa* activation. Consistent with this *Mitf* in mice is required for the transcription of many melanocyte genes, including tyrosinase-related genes with roles in melanocyte differentiation (Le Douarin and Kalcheim, 1999; (Kelsh et al., 2000b). Downstream of *nac/mitfa* there are other genes with important roles such as *spa/c-kit* (Parichy et al., 1999b) which is needed for melanophore dispersal and survival (Kelsh et al., 2000b; Parichy et al., 1999b). *dct* expression patterns in *spa* $^{-/-}$ embryos demonstrate that melanoblasts are present in normal numbers (Kelsh et al., 2000b). We shall use this knowledge on melanophore development to hypothesise the role of *shd/ALK* in iridophore development.

Although data is limited, ours and others, results to date do not deny the chromatoblast theory proposed by (Bagnara et al., 1979) where all pigment cell types have a common progenitor the chromatoblast. We describe a model of iridophore development incorporating this concept (Figures 5.11 and 5.12). The chromatoblast cell generates each of the three pigment cell-types, perhaps by the action of three different early specification genes. Based on results reported here and elsewhere we propose that specification occurs progressively via a melano-iridoblast. The absence of *shd/ALK* expression in neural crest cells of *cls*$^{-/-}$ embryos is consistent with *sox10* function being required for development of the chromatoblast.

Evidence for a common precursor for zebrafish melanoblasts and iridoblasts arises from the work done by Lister et al. (1999) in the *nac* mutant (*mitfa*). This is based on the fact that *nac*$^{-/-}$ mutants lack melanophores and their precursors but show an increased number of iridophores. These authors went further to explain the iridophore phenotype in *nac*$^{-/-}$ and suggested that another transcription factor *FoxD3* might be repressing *nac/mitfa* in this bipotent precursor cell to allow the iridophore lineage to develop (J. Lister personal communication). The role of *foxD3* in repressing melanogenesis thereby allowing other neural crest derivatives to differentiate has been shown in avian embryos before (Kos et al., 2001). *foxD3* morpholinos had no effect on iridophore number in *nac*$^{-/-}$ embryos suggesting the increase in iridophore number results from the absence of *nac* function (J. Lister personal communication). On the other hand, it was demonstrated that the same *foxD3* morpholino injected into *spa*$^{-/-}$ mutants (which have reduced numbers of melanophores and normal numbers of iridophores (Parichy et al., 1999a) decreases the number of iridophores and increases melanophore number (J.Lister personal communication). These experiments point a) to the existence of a common precursor for iridoblasts and melanoblasts and b) to *foxD3* having an inhibiting effect on *nac/mitfa* allowing iridoblasts to develop.

In the current work, it was shown that *foxD3* seems to be expressed by iridoblasts at 30 and 49 hpf. It is therefore suggested that at 30 hpf those cells might be the common precursor proposed by Lister (1999) where *foxD3* is simply shutting down *nac/mitfa* function and by doing so, allowing them to be specified for the first time as iridoblasts. A very important fact is that homozygous mutants for the *foxD3* gene, named *mother superior*, have a severe reduction of iridophores (Ella Knapik personal communication) confirming that this gene is likely to have a role in iridophore development in addition to its role as a marker for pre-migratory neural crest and glia of the PNS (Kelsh et al., 2000a) (see Figure 5.12).

The fact that we have found no differences between wild-type and *shd*$^{-/-}$ embryos in *foxD3* expression studies means that this gene is acting independently of the *shd/ALK* function. This result was surprising because it was telling us that even in the absence of *shd/ALK* gene function there are still cells that behave as if they were specified to the iridophore fate. This finding, if proved to be correct, is revealing that *foxD3* is acting upstream of *shd/ALK.* On the other hand, the fact that *foxD3* is still expressed at 49 hpf in these iridoblasts is intriguing as its inhibitory function would be expected to have finished by then. Perhaps *foxD3* has some other early role in the iridophore lineage, also prior to the *shd/ALK* function. Conceivably, it might be actively involved in *shd/ALK* recruitment as the next step in iridophore specification. Alternatively, may be *shd/ALK* is responsible for iridoblast survival or even for a very early differentiation step in iridoblasts that have already been specified by *foxD3.* It would be interesting to follow these links in the future (see Figure 5.12).

One other intriguing result presented here was the delayed expression of iridoblast markers in *nac*$^{-/-}$ embryos. The observation of an increased number of iridophores in *nac*$^{-/-}$ embryos would predict that many iridoblasts should be identified at 30 hpf using *shd/ALK* as a marker. This was not the case and only later could these cells be detected. The same applied to *ednrb1* expression that was detected later than 56hpf as opposed to 48 hpf in a wild-type situation. We suggested that in *nac*$^{-/-}$ embryos there was approximately a 10 hour delay in expressing these two markers. Therefore, a delay in the expression of iridoblast markers indicates that the lack of a functional Mitfa does not immediately result in the presence of more iridoblasts by default. In stead, it seams to take longer for such cells to form. One explanation would be that iridoblasts to become identifiable by *shd/ALK* and *ednrb1* markers need to have interacted with differentiated melanophores around 30 hpf. Perhaps because they need some factor secreted by melanophores. In the absence of this cell type as in the *nac*$^{-/-}$ situation, maybe the shd/ALK receptor has to be activated by a

different process that takes longer or has to wait for the same secreted factor to be generated by other cells types.

Coming back to the model and having added the early action of *foxD3* in the iridophore lineage, we need to include *shd/ALK* as a late specification gene, an iridoblast survival gene, or an early differentiation gene in iridophore development. The early differentiation role places *shd/ALK* in the same step as *dct* for the melanophore lineage development (Kelsh et al., 2000b). However, the *Dct* phenotype in mice is very different from the *shd* $^{-/-}$ phenotype, it shows a reduction in melanin and therefore the presence of paler melanocytes as opposed to absence of these cells (Le Douarin and Kalcheim, 1999). We therefore tend to discard this possibility.

An argument against *shd/ALK* being involved in survival or proliferation, in spite of having been assigned with this role in human tumours, is again based on the observation that the *shd* $^{-/-}$ phenotype is very strong. Unlike for example that of *spa*$^{-/-}$ embryos, which encode an RTK (c-kit) with proliferation/survival activity, where even the strongest phenotypes are never completely depleted of melanophores. On the other hand, we know that both in zebrafish and in mice the requirement for *c-kit* has been suggested to be after melanoblast proliferation (reviewed by Le Douarin and Kalcheim, 1999; Kelsh and Eisen, 2000). Therefore, an argument in favour of *shd/ALK* having a survival role is still possible if this function is vital for iridoblast survival before they proliferate. So that, in a strong phenotype we can not detect any iridoblasts.

The possibility of *shd/ALK* being involved in iridophore specification implies that it must be expressed in other cells apart from iridoblasts. At the moment we have no evidence that that is the case because we did not perform any double in situ hybridisations. However, the number, position and pattern of *shd/ALK* expressing cells strongly suggest that this gene is specifically expressed by iridoblasts from 30 hpf. Futhermore, the absence of

shd/ALK expression in *shd*$^{-/-}$ embryos supports this prediction. Consequently, it is plausible that *shd/ALK* expression is a marker of specification rather than a specifying sign. For this reason we shall consider both these scenarios from now on.

shd/ALK seems to start being down-regulated in the actual iridophore cell, again supporting that this gene is probably not involved in late iridophore maintenance or survival. In fact, such a role might be performed by *ednrb1* which is expressed by iridophores at later stages. Without the function of *ros/ednrb1,* adult iridophores are absent as shown by the *ros*$^{-/-}$ adult phenotype (Parichy et al., 2000a). A similar role in survival has been shown to be performed by *spa/c-kit* function in embryonic melanophores (Parichy et al., 1999b). However, the evidence for *ednrb1* having a survival role in embryonic iridophores is poor and we must remember that other genes will probably be identified for the iridophore lineage which might be better candidates for this function.

In addition to this speculative late function of *ednrb1* we question whether it might also have an early role in iridoblasts or indeed in common precursor pigment cells. Evidence for this is not robust as it arises solely from expression patterns. Nevertheless, expression of this gene between wild-type and *shd* $^{-/-}$ embryos is detectable as early as 24 hpf. At 30 hpf expression of *ednrb1* was detected in melanising cells too. According to Parichy (2000) this gene is expressed by all three pigment cell precursors until the time that each of them becomes differentiated. This indicates that at an early stage, possibly before melanisation, the cells expressing *ednrb1* might be the common precursors for all the pigment cell lineages proposed by Bagnara et al. (1979).

The fact that *ednrb1* expression ceases with differentiation of the different cell types makes us speculate that perhaps this gene has a role in keeping cells away from terminal differentiation during migration. Similar function was reported for this gene in the mouse (Shin, 1999), where the Ednrb/Edn3 system has been proposed to be involved in retaining

an undifferentiated status of neural crest cells so that these undergo migration through the whole gut. May be this function is conserved in fish pigment cells, so that by indirectly regulating the timing of migration they become progressively differentiated, first in melanophores and xanthophores and at last in iridophores.

Finally, for the model to be complete, genes involved in the differentiation of the actual reflective platelets would need to be identified. Candidates for these genes would be the genes causing dull iridophore phenotypes isolated in the Tuebingen screen in 1996, such as *brassy, punkt, pepita or pile-up* (Kelsh et al., 1996).

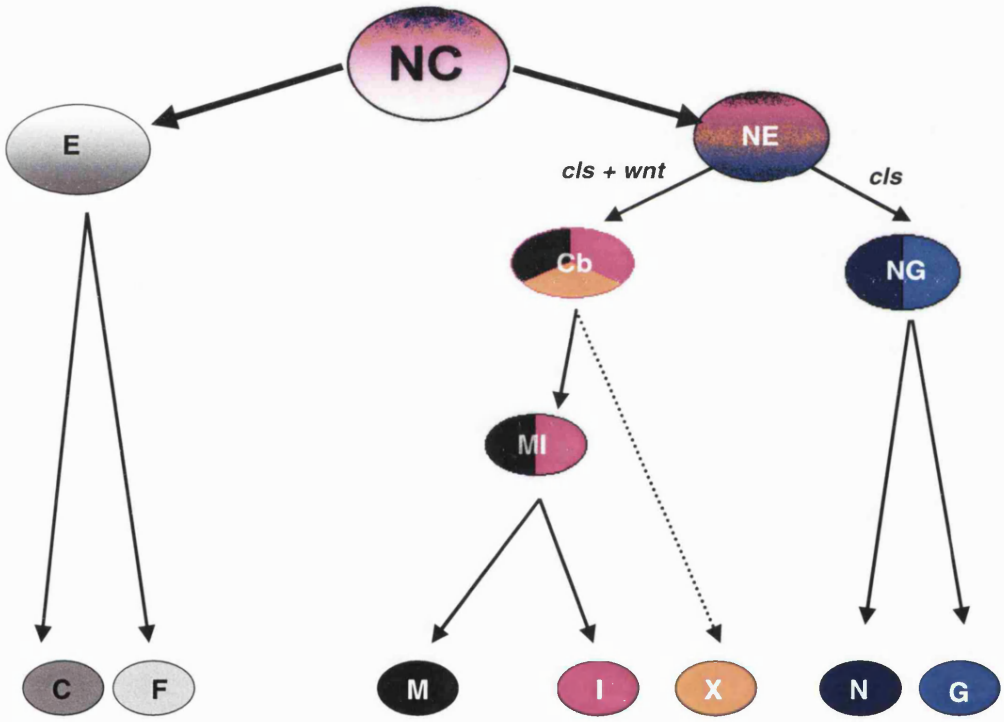


Figure 5.11 Progressive fate restriction of neural crest cells in zebrafish. Adapted from (Kelsh and Raible, 2002). A neural crest cell (NC) gives rise to ectomesenchymal (E) and non-ectomesenchymal (NE) progenitors. The former cells generate craniofacial cartilage (C) and fin mesenchyme (F) amongst other fates. Non-ectomesenchymal precursors are initially pluripotent and *cls* gene function is needed for all its different derivatives. It is suggested here that there might be a common pigment cell precursor known as a chromatoblast (Cb) and a common neuron-glial precursor (NG) which will give rise to neurons (N) and glia (G). Wnt signaling and *sox10* are implicated in generating the pigment cell lineage. The chromatoblast (Cb) theory was suggested by Bagnara et al.(1979). It is possible that a chromatoblast then becomes fate restricted to a Melano-iridoblast (MI) that in turn through the action of specific genes gives rise to melanophores (M) and iridophores (I) (see Figure 5.12 for close up of this region with detailed explanation). At the moment little is known about xanthophore (X) development.

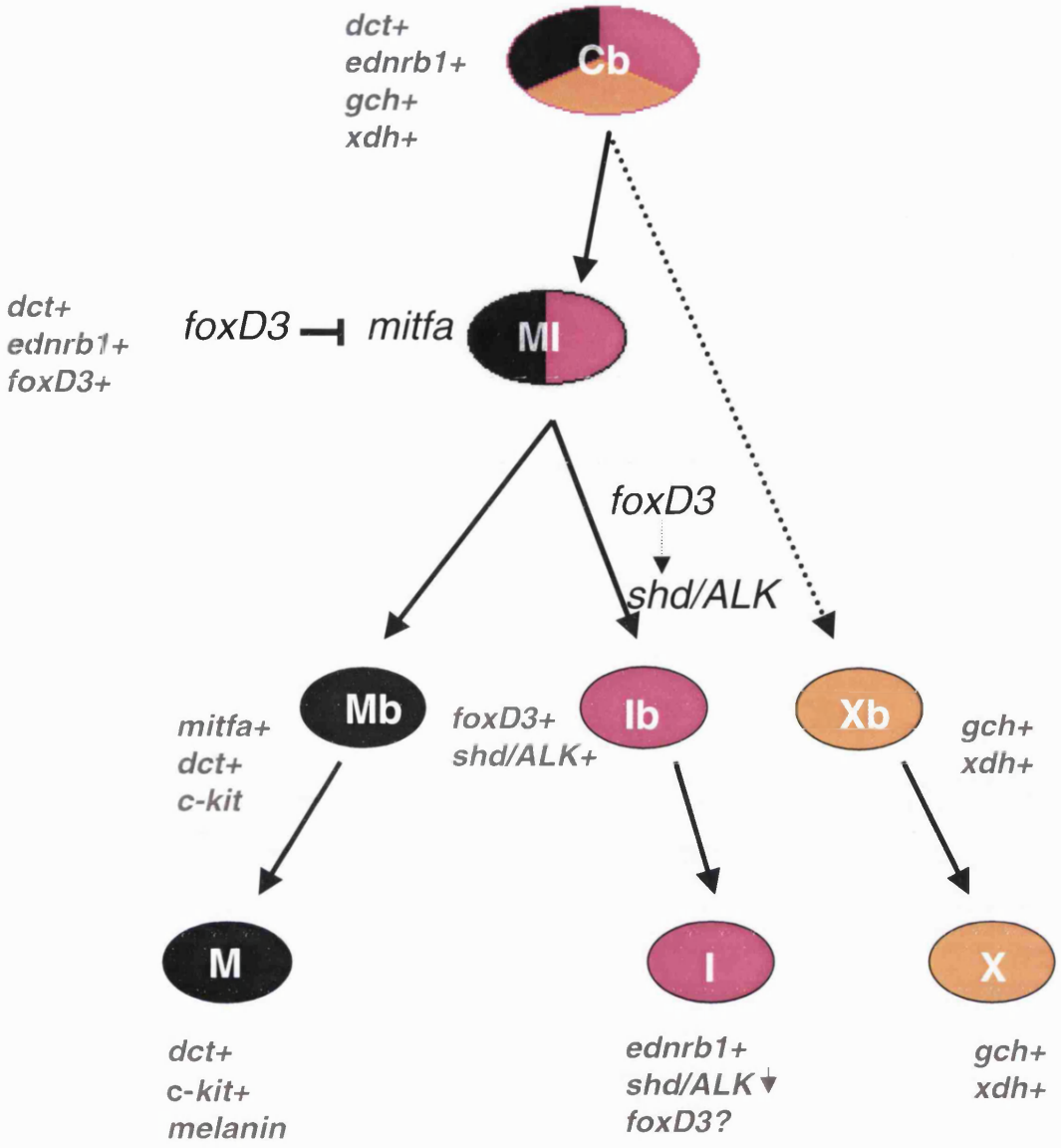


Figure 5.12 Model for the common pigment cell precursor. The chromatoblast (Cb) hypothesis is supported by *ednrb1* expression in cells that co-express specific markers for each pigment cell type (expression of *shd/ALK* to be confirmed). It is possible that a chromatoblast then becomes fate restricted to a melano-iridoblast (MI) where the transcriptional repressor *foxD3* inhibits *mitfa* during a period of time to allow iridoblasts (Ib) to develop. Normally, *mitfa* would continue to form melanoblasts (Mb), which express both *mitfa* and *dct,* and would fully differentiate into melanised cells with the help of *dct* amongst other genes. Then, melanophores would need *c-Kit* function to disperse properly and survive. For the iridophore lineage to form the function of *shd/ALK* is necessary but perhaps not sufficient as *foxD3* expression in late iridoblasts suggests an additional unknown role for this gene in these cells. It is conceivable that *foxd3* might be directly or indirectly transcriptionaly activating *shd/ALK* as a specification step. Differentiated iridophores (I) express *ednrb1* perhaps involved in maintenance and survival of these cells. Iridophores also express *shd/ALK* at lower levels indicating down-regulation of this gene. It is questioned whether iridophores might still express *foxD3*. Much less is known about the xanthophore lineage (X) at the current time, except that in a *cls*$^{-/-}$ embryo these are the only cells that develop (although in lower numbers) perhaps because xanthophores are a default state for pigment cells. However, we can not rule out the possibility of the presence of a Melano-xanthoblast. Xanthoblasts (Xb) and xanthophores express *gch* and *xdh*, two markers of xanthophore differentiation. Genes expressed by the different cell types are in grey.

Chapter 6

General Discussion and conclusions

As part of a positional cloning project the starting point of this work was to map *shd* to one of the zebrafish 25 linkage groups. The region of previously mapped genes *otx1* and *id2* on linkage group 17 was closest to the markers nearest *shd* . Synteny between this region and human chromosome 2p13-2p25 was confirmed for four zebrafish genes which suggested that a human homologue of *shd* would likely lie within this region of chromosome 2p. Isolation of genomic clones and application of combined restriction analysis, phenotypic rescue by genomic DNA micro-injection and sequencing narrowed down the genomic interval containing *shd* to a region where only one candidate gene could be identified. This predicted candidate gene was an RTK similar to both human ALK and LTK genes. One of them, ALK, had indeed been mapped to a position within the region of human chromosome 2p13-2p25. However, human LTK on chromosome 15 is very close to both IVD and ITPKA, genes orthologous to zebrafish predicted *ivd* and *itpka* that were found in our genomic contig (PACs 1 and 3, respectively). It was therefore concluded that in terms of synteny both these RTKs could be candidates for *shd*. It was only after a detailed phylogenetic analysis (in collaboration with Professor Laurence Hurst, University of Bath) that we could conclude that *shd* is the true orthologue of the mammalian *ALK* genes.

cDNA cloning and expression analysis of the zebrafish ALK/LTK gene strengthened its candidacy since: a) its expression pattern included iridophores and their precursors and b) a morpholino designed to this RTK successfully phenocopied the *shd* mutant phenotype.

Two different cDNA clones were identified, these were identical except for the presence of 78 nucleotides encoding an extra 26 amino acids in clone 1. The difference between clones appeared not to be caused by a differential splice event since consensus splice sites were absent in the genomic sequence. It seemed more likely therefore, to result from a deletion during the cloning procedure.

Mammalian ALK proteins showed much wider similarity to zebrafish RTK predicted ORF in both the extracellular and intracellular region than did LTK proteins. Thus, structurally, the zebrafish RTK was more reminiscent of an ALK protein than an LTK. In particular, the MAM domains identified in the zebrafish RTK are only present in ALK proteins. Again, it was only through the phylogenetic analysis based on the most conserved domain (tyrosine kinase domain) between zebrafish, mouse, rat and human ALKs and LTKs, both at nucleotide and amino acid levels, that we confirmed that our RTK was indeed an ALK orthologue.

The comparison of expression patterns between zebrafish RTK and the two mammalian genes was not informative because mammals lack iridophores. The only pigment cell type present in mammals is the melanocyte (melanophore equivalent in a homeothermic organism) but there are no reports of ALK or even LTK expression in these cells.

Antibodies raised against the ALK ligand binding domain (within one of the mouse MAM domains) inhibit PTN and MK from binding to mouse ALK (Stoica et al., 2001; Stoica et al., 2002). We thus know that both PTN and MK are potential ligands for the mammalian ALK receptor (Stoica et al., 2001; Stoica et al., 2002). There is evidence that these proteins in higher vertebrates are involved in a wide range of activities. MK has a neurotrophic function as it promotes outgrowth and survival of cultivated neurons (Michikawa et al., 1993), its up-regulation has been observed in a number of tumours (including those involving ALK), where it seems to have a role in invasion and survival of tumour cells (Stoica et al., 2002). Furthermore, it has been reported in angiogenesis, wound healing and tissue repair processes (Muramatsu, 2002; Wada et al., 2002; Yoshida et al., 1995). For instance, MK expression (which is normally very restricted in the adult) re-emerges in the adult brain during pathologic alterations such as Alzheimer's disease or cerebral infarction (Yoshida et al., 1995). PTN has also been implicated in endothelial cell proliferation and angiogenesis and it has been

demonstrated to function as a tumour growth factor (Muramatsu, 2002; Stoica et al., 2001; Stoica et al., 2002).

There are two paralogous *midkine* (*mdk*) genes recently identified in zebrafish and one EST for the *pleiotrophin* (*ptn*) gene (Winkler et al., 2003). Therefore, zebrafish appears to have orthologues of both the tetrapod genes that have products implicated in ALK signalling. One of the fish midkines, *mdkb*, has been suggested to act as a planar signal in early neural patterning. Its later expression in the dorsal neural tube together with the finding that ectopic *mdkb* promotes neural crest cell formation (Winkler and Moon, 2001) makes this gene the best candidate for encoding a zebrafish shd/ALK ligand.

Later expression of zebrafish *mdkb* or indeed of the *ptn* EST have not been reported. These expression studies would be critical in revealing whether *mdkb* and/or *ptn* have the right localisation during iridophore development to be seriously considered as encoding candidates for shd/ALK ligands. Therefore, we strongly suggest such studies should constitute the next step in the line of research of this project.

We would like to end this thesis by proposing a model for the likely role of *shd/ALK* in the iridophore pigment cell lineage. We begun these studies considering *shd* to be involved in an early step in iridophore development, perhaps fate specification (see Chapter 1). However, by the end of this work we propose at least three possible roles for *shd* (models A, B and C), based on the following key observations:

Model A- shd/ALK as a gene needed for iridophore specification

This model considers that the *foxD3* expressing cells described in Chapter 5 are bipotent melano-iridoblast cells. These bipotent cells would reach their presumptive iridophore locations in the embryonic zebrafish and wait for further instructions on whether to follow the iridophore lineage. Because *foxD3* positive cells are present at 30 and 49 hpf both in wild-types and *shd* $^{-/-}$ in numbers and locations that are typical of final destinations for iridophores, it may reflect that this gene has an early role in iridophore specification before the *shd/ALK* function. Hence we believe the presumptive function of *foxD3* in these cells is to enable the bipotent cells to become iridoblasts by repressing the early melanoblast specifying gene *mitfa.* The role of *shd/ALK* could therefore be one of specification, perhaps directly or indirectly activated by *foxD3* (model A in Figure 6.1). The specification hypothesis assumes that *shd/ALK* is necessary for restricting the fate of the bipotent melano-iridoblast to iridoblast. Consequently, in a *shd* $^{-/-}$ environment the *foxD3* expressing cells that we saw would be this bipotent precursor waiting for the final fate restriction step. What happens to these cells is worth investigation. Thus, we propose to look for cell death at around 49-50 hpf. On the other hand, in a wild-type environment we predict that *foxD3* would be co-expressed with *shd/ALK.*

To assign a definite role for *shd/ALK* in iridophore development it would be crucial to understand the outcome of signalling through this RTK. Our predictions, in case *shd/ALK* is the specifying gene, are that the signalling through the shd/ALK receptor will result in the transcription of early differentiation iridophore genes responsible, for example, for the formation of guanine platelets (model A in Figure 6.1).

Model B- shd/ALK as a gene needed for iridoblast survival

Alternatively, the specification role might be performed by *foxD3* as a result of inhibiting *mitfa* and *shd/ALK* may be involved in early survival of iridoblasts. Therefore the survival hypothesis (model B in Figure 6.1) assumes that the 'bipotent' cell is already specified to an iridoblast by *foxD3* action, perhaps in conjunction with other genes. As a result, in a *shd*$^{-/-}$ environment, the *foxD3* expressing cells would almost certainly die whilst in a wild-type environment we also expect these cells to co-express *foxD3* and *shd/ALK*. Whether *foxD3* is maintained in iridophores or is down regulated after 49 hpf has to be further investigated. We propose to test for apoptosis by TUNEL and follow the fate of these cells by single cell iontophoretic labeling.

In case *shd/ALK* is a survival gene (model B in Figure 6.1), we predict that the signalling outcome might result in the activation of the PI3-kinase axis resulting in phosphorylation of the downstream target molecule Akt, an anti-apoptotic effector of PI3-kinase signalling (Brunet et al., 1999). This cascade of events has been reported in human cell-lines caused by activation of ALK by MK and PTN binding (Stoica et al., 2002). We propose that the same signal transduction pathway may be needed for iridoblast survival. The ability of Akt to promote cell survival by inhibiting transcription of genes that belong to the apoptotic machinery is well documented (Brunet et al., 1999; Brunet et al., 2001a; Brunet et al., 2001b; Datta et al., 1999).

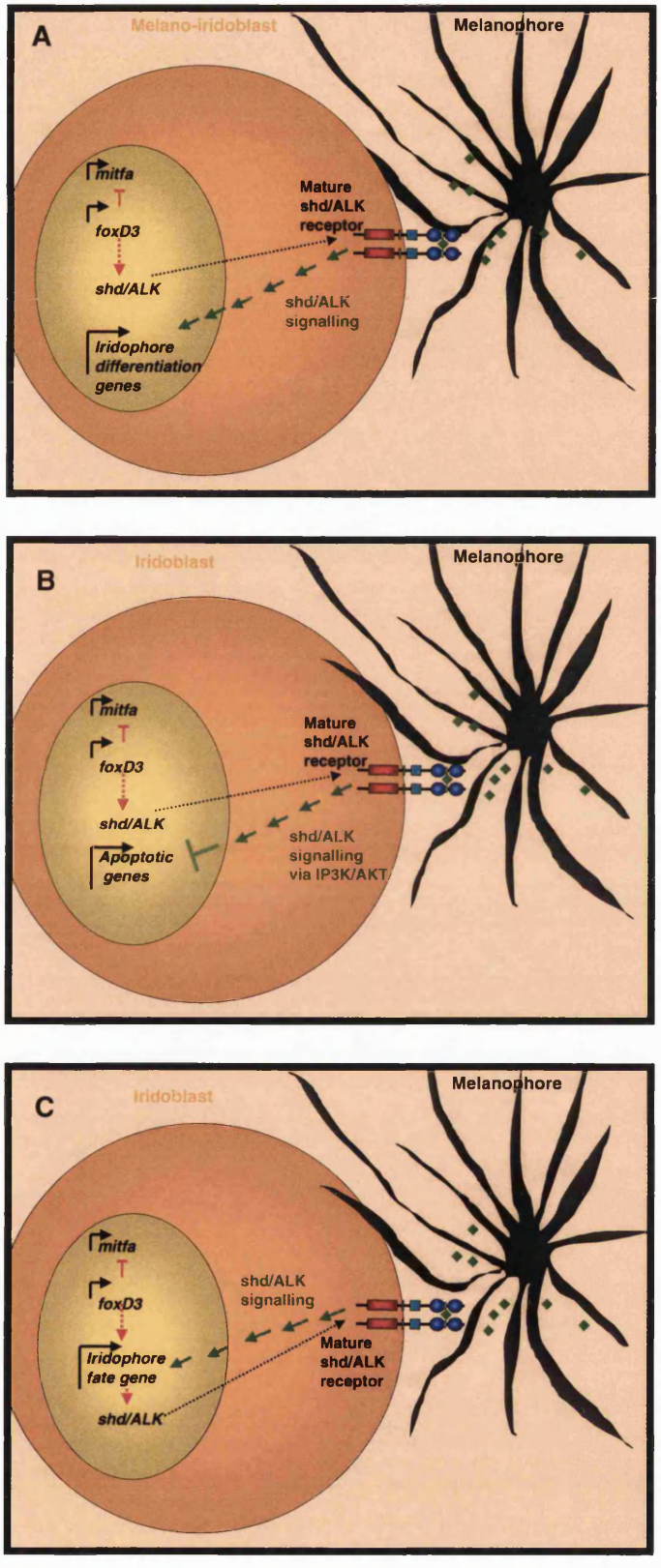
Model C- shd/ALK as a gene needed for iridoblast maintenance

Other possibility, (model C in Figure 6.1) is that *foxD3* while repressing *mitfa*, activates an iridophore fate gene, which in turn activates the transcription of the *shd/ALK* gene. The signalling through the mature shd/ALK receptor could then result in the regulation of the upstream iridophore fate gene, by a positive feedback loop. Consequently, the function of

shd/ALK would be maintenance of the iridoblast fate. This model is consistent with our observation that *shd/ALK* expression is absent in *shd*$^{-/-}$ mutants, because in this context shd/ALK signalling would be needed to maintain *shd/ALK* gene expression. We would expect to find apoptotic iridoblasts in the absence of shd/ALK function.

A similar regulatory model has been suggested for c-kit signalling (Hemesath et al., 1998). These authors have shown that c-kit signalling in vitro activates the ERK/MAP kinase pathway, which in turn was able to phosphorylate Mitf in melanocytes The ability of Mitf to transactivate was successfully tested on a luciferase reporter driven by the tyrosinase promoter (Hemesath et al., 1998). We suggest that shd/ALK signalling might also use a MAP kinase pathway for targeting the phosphorylation of the putative iridophore fate gene because in addition to the reported IP3 kinase pathway, human ALK has also been shown to activate the MAP kinase transduction pathway (Stoica et al., 2002).

The activation of the shd/ALK receptor by the putative growth factors mdkb or ptn assumes that in the environment around this receptor there must be growth factor secreting cells. We suggest that perhaps a close source for these factors might be the melanophore cells, since iridophores always appear somehow associated with melanophores (see wild-type phenotype in Chapter 5). In fact, because iridophores fully differentiate around 20 hours after melanophores it means that melanophores could secrete the growth factors that activate the shd/ALK receptor in their neighbouring melano-iridoblasts/iridoblasts (see Figure 6.1). This hypothesis would also explain why in a *nac*$^{-/-}$embryo (which lacks all melanophores) we see a delay in iridoblast development (Chapter 5). It is conceivable that melanophores are the closest but not the only source of these growth factors explaining why iridophores are still formed, although later than normal.


A
Melano-iridoblast
Melanophore
mitfa
foxD3
shd/ALK
Iridophore differentiation genes
Mature shd/ALK receptor
shd/ALK signalling
B
Iridoblast
Melanophore
mitfa
foxD3
shd/ALK
Apoptotic genes
Mature shd/ALK receptor
shd/ALK signalling via IP3K/AKT
C
Iridoblast
Melanophore
mitfa
foxD3
Iridophore fate gene
shd/ALK
shd/ALK signalling
Mature shd/ALK receptor

In summary, we have positionally cloned *shd* and shown it encodes an RTK orthologue to the mammalian *ALK* gene which has a poorly understood normal function, in contrast to its better known role in human cancer. The phenotypic observations on zebrafish *shd* mutants and other iridophore mutants compared to their wild-type siblings allowed us to generate models in which we suggest that *shd/ALK* plays an important role in either iridoblast specification, survival or maintenance. We suggest further experiments for testing these roles as well as for testing the candidacy of the ligands for the zebrafish shd/ALK receptor.

Figure 6.1 Model for the specification activity of the *shd/ALK* gene. (**A**) We propose that in the bipotent melano-iridoblast cell the transcription factor *foxD3* inhibits the transcription of *mitfa* and thereby allows the iridophore lineage to develop. *foxD3* is also activating the transcription of *shd/ALK* either directly or indirectly (dotted red arrow). The mature shd/ALK receptor is then activated by the binding of putative growth factors (green diamonds), such as mdkb and/or ptn. We speculate that these factors might be secreted by melanophores amongst other cells. Ligand binding results in receptor dimerisation and consequent RTK intracellular signalling (green arrows). We suggest this cascade of events has the final objective of activating the transcription of iridophore differentiation genes that will be responsible for the formation of shiny iridophores by, for example producing the platelets made of stacks of guanine crystals. (**B**) Alternatively, the role of *shd/ALK* gene is needed for iridoblast survival whereby shd/ALK signalling activates the IP3K/AKT pathway that is known to inhibit the transcription of apoptotic genes and may therefore mediate iridoblast cell survival. (**C**) A third possibility would consist in shd/ALK signalling result in the transcriptional regulation of an iridophore fate gene, downstream of foxD3, by a positive feedback loop. This last model would be one of iridoblast fate maintenance.

APPENDIX

1. CLUSTAL W (1.82) multiple sequence alignment between wild-type clones 1 to 3. Sequencing primers are annotated in pink. The 1200bp fragment used for to the riboprobe is annotated with blue stars with respective primers in blue. Long distance PCR primers are in green and RACE PCR primers are in light blue. Start and Stop codons are boxed in red. Morpholino sequence is in bold.

```
clone1wt        GGTGACAGTGCAGTACTCTGCCTGTAGCAGTCATGAAGTTGGGTTTCTGGCATTCGATTC 720
clone3wt        GGTGACAGTGCAGTACTCTGCCTGTAGCAGTCATGAAGTTGGGTTTCTGGCATTCGATTC 720
clone2WT        GGTGACAGTGCAGTACTCTGCCTGTAGCAGTCATGAAGTTGGGTTTCTGGCATTCGATTC 720
                ************************************************************

clone1wt        ACTGGAGCTGAAGAACTGTGTCATGGGTGATGATTATGTGGATTTGGGTTCCGACTGTGA 780
clone3wt        ACTGGAGCTGAAGAACTGTGTCATGGGTGATGATTATGTGGATTTGGGTTCCGACTGTGA 780
clone2WT        ACTGGAGCTGAAGAACTGTGTCATGGGTGATGATTATGTGGATTTGGGTTCCGACTGTGA 780
                ************************************************************

clone1wt        GAAATACTCATCTCTTCAATGTCATTCTGGAGGCTGCATTGAGAAACAAAGAGTGTGTGA 840
clone3wt        GAAATACTCATCTCTTCAATGTCATTCTGGAGGCTGCATTGAGAAACAAAGAGTGTGTGA 840
clone2WT        GAAATACTCATCTCTTCAATGTCATTCTGGAGGCTGCATTGAGAAACAAAGAGTGTGTGA 840
                ************************************************************

clone1wt        TTTCCACACTGACTGCCCAGAAGGGGAGGACGAGGGCTTAATATGCAGCACTTTACCTTT 900
clone3wt        TTTCCACACTGACTGCCCAGAAGGGGAGGACGAGGGCTTAATATGCAGCACTTTACCTTT 900
clone2WT        TTTCCACACTGACTGCCCAGAAGGGGAGGACGAGGGCTTAATATGCAGCACTTTACCTTT 900
                ************************************************************

clone1wt        GGGTTCATACTGTTCATTTGAGTTGGGCTCCTGTGGTTGGTCAGCGGCTGACACACAGTC 960
clone3wt        GGGTTCATACTGTTCATTTGAGTTGGGCTCCTGTGGTTGGTCAGCGGCTGACACACAGTC 960
clone2WT        GGGTTCATACTGTTCATTTGAGTTGGGCTCCTGTGGTTGGTCAGCGGCTGACACACAGTC 960
                ************************************************************

clone1wt        TTCTTGGAGACTAGTTAGTGGACAACAGCTGATTGAAGACACACACCTGCTGGGCACAAC 1020
clone3wt        TTCTTGGAGACTAGTTAGTGGACAACAGCTGATTGAAGACACACACCTGCTGGGCACAAC 1020
clone2WT        TTCTTGGAGACTAGTTAGTGGACAACAGCTGATTGAAGACACACACCTGCTGGGCACAAC 1020
                ************************************************************

                                                          T7-2 ──►
clone1wt        TCTGAAGAACACTCAAGGGCATTTCCTTTTCTTGAAGGTCAGAGGTCACGGTGATGAAAG 1080
clone3wt        TCTGAAGAACACTCAAGGGCATTTCCTTTTCTTGAAGGTCAGAGGTCACGGTGATGAAAG 1080
clone2WT        TCTGAAGAACACTCAAGGGCATTTCCTTTTCTTGAAGGTCAGAGGTCACGGTGATGAAAG 1080
                ********************************************** ◄──
                                                          T7-2RC

clone1wt        AGAGGCGCTGGTTCAAAGCCCTGCATTGCCTTCAACTATATCCAATCAGGACTGTCAGTT 1140
clone3wt        AGAGGCGCTGGTTCAAAGCCCTGCATTGCCTTCAACTATATCCAATCAGGACTGTCAGTT 1140
clone2WT        AGAGGCGCTGGTTCAAAGCCCTGCATTGCCTTCAACTATATCCAATCAGGACTGTCAGTT 1140
                     ********************************************************

clone1wt        GCAGTTTTCTCTCTACCGGTATGGGGATTTCAATGGAACAGTGCTTTTGTCTGTGGTGGA 1200
clone3wt        GCAGTTTTCTCTCTACCGGTATGGGGATTTCAATGGAACAGTGCTTTTGTCTGTGGTGGA 1200
clone2WT        GCAGTTTTCTCTCTACCGGTATGGGGATTTCAATGGAACAGTGCTTTTGTCTGTGGTGGA 1200
                ************************************************************

clone1wt        GAGCGGAGCCTCAGCACCAGCACTGATTTGGGAGAGATCCGGACACTGGAAAGACGCCTG 1260
clone3wt        GAGCGGAGCCTCAGCACCAGCACTGATTTGGGAGAGATCCGGACACTGGAAAGACGCCTG 1260
clone2WT        GAGCGGAGCCTCAGCACCAGCACTGATTTGGGAGAGATCCGGACACTGGAAAGACGCCTG 1260
                ************************************************************

clone1wt        GCAAGAGATCACTCTGCCGATCACAGAGATTTTAAATGGCTTCCATCTTAAAGTGCAGGC 1320
clone3wt        GCAAGAGATCACTCTGCCGATCACAGAGATTTTAAATGGCTTCCATCTTAAAGTGCAGGC 1320
clone2WT        GCAAGAGATCACTCTGCCGATCACAGAGATTTTAAATGGCTTCCATCTTAAAGTGCAGGC 1320
                ************************************************************

clone1wt        TTTCTGGACTTCTGGCTCCAAGGCTGACATTGCACTCGATGACATTTCATTAAGTGCAGC 1380
clone3wt        TTTCTGGACTTCTGGCTCCAAGGCTGACATTGCACTCGATGACATTTCATTAAGTGCAGC 1380
clone2WT        TTTCTGGACTTCTGGCTCCAAGGCTGACATTGCACTCGATGACATTTCATTAAGTGCAGC 1380
                ************************************************************

clone1wt        ATGTTTTGACACAGAACTGAATGAACTGCTTCATGAAGGACTACCGCATGATTTAGACTT 1440
clone3wt        ATGTTTTGACACAGAACTGAATGAACTGCTTCATGAAGGACTACCGCATGATTTAGACTT 1440
clone2WT        ATGTTTTGACACAGAACTGAATGAACTGCTTCATGAAGGACTACCGCATGATTTAGACTT 1440
                ************************************************************
```


clone1wt CAGTCCTCTTCCAGAACCATCAGCATCAGAGGCGTCTCCAATTACATGGTGGTTTACATC 1500
clone3wt CAGTCCTCTTCCAGAACCATCAGCATCAGAGGCGTCTCCAATTACATGGTGGTTTACATC 1500
clone2WT CAGTCCTCTTCCAGAACCATCAGCATCAGAGGCGTCTCCAATTACATGGTGGTTTACATC 1500
Forward shd1
clone1wt CTGTGGAGCGAGTGGACCTTTCGGCCCGACTCAAGCCCAGTGTGACAGTGCCTACCGAAA 1560
clone3wt CTGTGGAGCGAGTGGACCTTTCGGCCCGACTCAAGCCCAGTGTGACAGTGCCTACCGAAA 1560
clone2WT CTGTGGAGCGAGTGGACCTTTCGGCCCGACTCAAGCCCAGTGTGACAGTGCCTACCGAAA 1560
clone1wt CACCAATGTCAGCGTTGTGGTGGGGAAAGAGGGTCCTCTCAGGGGTGTACAGATGTGGAA 1620
clone3wt CACCAATGTCAGCGTTGTGGTGGGGAAAGAGGGTCCTCTCAGGGGTGTACAGATGTGGAA 1620
clone2WT CACCAATGTCAGCGTTGTGGTGGGGAAAGAGGGTCCTCTCAGGGGTGTACAGATGTGGAA 1620
clone1wt GGTCCCTGCTACCAACACGTACAAGATTTCAGCCTATGGAGCTGCAGGAGGCAAAGGAGC 1680
clone3wt GGTCCCTGCTACCAACACGTACAAGATTTCAGCCTATGGAGCTGCAGGAGGCAAAGGAGC 1680
clone2WT GGTCCCTGCTACCAACACGTACAAGATTTCAGCCTATGGAGCTGCAGGAGGCAAAGGAGC 1680
T7-3
clone1wt AAAGAACCATAACAAGCGCTCACACGGTGTGTTCATCTCAGCCACCTTCCCCCTGGAGAA 1740
clone3wt AAAGAACCATAACAAGCGCTCACACGGTGTGTTCATCTCAGCCACCTTCCCCCTGGAGAA 1740
clone2WT AAAGAACCATAACAAGCGCTCACACGGTGTGTTCATCTCAGCCACCTTCCCCCTGGAGAA 1740
T7-3RC
clone1wt AGGTGACATCCTCTACATCTTAATTGGCCACCAAGGAGAAGATGCCTGTCCAGGAAGGAA 1800
clone3wt AGGTGACATCCTCTACATCTTAATTGGCCACCAAGGAGAAGATGCCTGTCCAGGAAGGAA 1800
clone2WT AGGTGACATCCTCTACATCTTAATTGGCCACCAAGGAGAAGATGCCTGTCCAGGAAGGAA 1800
clone1wt CCCCCAAACCCATAAAATCTGTCTGGGCGAGTCGTCTGTGATTGAAGATGGTTTTGACAG 1860
clone3wt CCCCCAAACCCATAAAATCTGTCTGGGCGAGTCGTCTGTGATTGAAGATGGTTTTGACAG 1860
clone2WT CCCCCAAACCCATAAAATCTGTCTGGGCGAGTCGTCTGTGATTGAAGATGGTTTTGACAG 1860
clone1wt CGACGGTTCAGCCTTGAAGTGGGCAGGGGGTGGCGGCGGTGGAGGGGGAGCCACATATAT 1920
clone3wt CGACGGTTCAGCCTTGAAGTGGGCAGGGGGTGGCGGCGGTGGAGGGGGAGCCACATATAT 1920
clone2WT CGACGGTTCAGCCTTGAAGTGGGCAGGGGGTGGCGGCGGTGGAGGGGGAGCCACATATAT 1920
clone1wt ATATCGGATGGAGAACGGGCAACCACTTCCTTTGCTGATTGCAGCTGGAGGGGGAGGAAA 1980
clone3wt ATATCGGATGGAGAACGGGCAACCACTTCCTTTGCTGATTGCAGCTGGAGGGGGAGGAAA 1980
clone2WT ATATCGGATGGAGAACGGGCAACCACTTCCTTTGCTGATTGCAGCTGGAGGGGGAGGAAA 1980
clone1wt GGCCTACCTTGAGGATCCAGAAAGCAGCCAGGACCAGAGTTTTCGGGAACAGTACGAGAA 2040
clone3wt GGCCTACCTTGAGGATCCAGAAAGCAGCCAGGACCAGAGTTTTCGGGAACAGTACGAGAA 2040
clone2WT GGCCTACCTTGAGGATCCAGAAAGCAGCCAGGACCAGAGTTTTCGGGAACAGTACGAGAA 2040
clone1wt TGACACCACCGTTTCTGGTGTCAGTGGCAGATCTGGAGCAGCAGGTGGAGGAGGTGGATG 2100
clone3wt TGACACCACCGTTTCTGGTGTCAGTGGCAGATCTGGAGCAGCAGGTGGAGGAGGTGGATG 2100
clone2WT TGACACCACCGTTTCTGGTGTCAGTGGCAGATCTGGAGCAGCAGGTGGAGGAGGTGGATG 2100
clone1wt GAGCGACGTGTCCTCTCTCTCATGGGCAGGGAAATCTCTTGTTGAAGGGGGTCAAGGAGG 2160
clone3wt GAGCGACGTGTCCTCTCTCTCATGGGCAGGGAAATCTCTTGTTGAAGGGGGTCAAGGAGG 2160
clone2WT GAGCGACGTGTCCTCTCTCTCATGGGCAGGGAAATCTCTTGTTGAAGGGGGTCAAGGAGG 2160
T7-4
clone1wt CTCATCCTGTCCTGAGGCTTTGTCAGTGCTGGGATGGGCTACATTTGGAGGATTCGGAGG 2220
clone3wt CTCATCCTGTCCTGAGGCTTTGTCAGTGCTGGGATGGGCTACATTTGGAGGATTCGGAGG 2220
clone2WT CTCATCCTGTCCTGAGGCTTTGTCAGTGCTGGGATGGGCTACATTTGGAGGATTCGGAGG 2220
T7-4RC


clone1wt AGGAGGAGGCGCGTGCTCCGCTGGAGGGGGCGGTGGAGGATACAGAGGTGGTGATGCACC 2280
clone3wt AGGAGGAGGCGCGTGCTCCGCTGGAGGGGGCGGTGGAGGATACAGAGGTGGTGATGCACC 2280
clone2WT AGGAAGAGGCGCGTGCTCCGCTGGAGGGGGCGGTGGAGGATACAGAGGTGGTGATGCACC 2280
T7-4b
clone1wt TCTTCTTGACGACATCTCAGCAGATGGACAAGATGGCCTCTCTTTTGTTCACCCCATGGG 2340
clone3wt TCTTCTTGACGACATCTCAGCAGATGGACAAGATGGCCTCTCTTTTGTTCACCCCATGGG 2340
clone2WT TCTTCTTGACGACATCTCAGCAGATGGACAAGATGGCCTCTCTTTTGTTCACCCCATGGG 2340
clone1wt AAAGATATTTCTGCAACCTCTGGCAGCCATGGAGAGTCACGGCGAAGCTGAAATTGTGGT 2400
clone3wt AAAGATATTTCTGCAACCTCTGGCAGCCATGGAGAGTCACGGCGAAGCTGAAATTGTGGT 2400
clone2WT AAAGATATTTCTGCAACCTCTGGCAGCCATGGAGAGTCACGGCGAAGCTGAAATTGTGGT 2400
clone1wt CTATCTGAACTGCAGCCACTGTAAAACCCAGAGCTGCAAGCGCGATGAAGACACCAAGCT 2460
clone3wt CTATCTGAACTGCAGCCACTGTAAAACCCAGAGCTGCAAGCGCGATGAAGACACCAAGCT 2460
clone2WT CTATCTGAACTGCAGCCACTGTAAAACCCAGAGCTGCAAGCGCGATGAAGACACCAAGCT 2460
clone1wt CATCCTGTGCCTCTGCGACAGCGACGAGGTTCTGGCACCAGACAACGTCACCTGCGCAGG 2520
clone3wt CATCCTGTGCCTCTGCGACAGCGACGAGGTTCTGGCACCAGACAACGTCACCTGCGCAG- 2519
clone2WT CATCCTGTGCCTCTGCGACAGCGACGAGGTTCTGGCACCAGACAACGTCACCTGCGCAG- 2519
clone1wt TACCAAACACAGCCTCTGCCAATTTATCAACAAACATCTCCAACACAACTCCTCTCCCTT 2580
clone3wt
clone2WT
clone1wt GGTGTGTCCTCCTCTAGTGGTTCCTATGGGCTCTCTGGCAGACGGGCCTCCATCCCTGGT 2640
clone3wt TGGTTCCTATGGGCTCTCTGGCAGACGGGCCTCCATCCCTGGT 2562
clone2WT TGGTTCCTATGGGCTCTCTGGCAGACGGGCCTCCATCCCTGGT 2562
clone1wt CTTCATCATGGCAGTGATCGTGTCCACAGTGGTGACCGGTGTCGTTCTGACCTGTGCCAG 2700
clone3wt CTTCATCATGGCAGGGATCGTGTCCACAGTGGTGACCGGTGTCGTTCTGACCTGTGCCAG 2622
clone2WT CTTCATCATGGCAGGGATCGTGTCCACAGTGGTGACCGGTGTCGTTCTGACCTGTGCCAG 2622
clone1wt CCTGACTCTCATATATTATCGTAAGAAGAACCACCTGCATGCGGTCAGGATTCGACTGCA 2760
clone3wt CCTGACTCTCATATATTATCGTAAGAAGAACCACCTGCATGCGGTCAGGATTCGACTGCA 2682
clone2WT CCTGACTCTCATATATTATCGTAAGAAGAACCACCTGCATGCGGTCAGGATTCGACTGCA 2682
Reverse shd1
clone1wt GAGTCCAGAGTACAAGCTTAGCAAAATTCGCTCGTCCACCATCATGACCGACTACAACCC 2820
clone3wt GAGTCCAGAGTACAAGCTTAGCAAAATTCGCTCGTCCACCATCATGACCGACTACAACCC 2742
clone2WT GAGTCCAGAGTACAAGCTTAGCAAAATTCGCTCGTCCACCATCATGACCGACTACAACCC 2742
clone1wt CAACTATGGTTATTTCGGAAAGGCAGCCTCTCTGAGTGAACTGAAGGAAGTGCCGCGTAA 2880
clone3wt CAACTATGGTTATTTCGGAAAGGCAGCCTCTCTGAGTGAACTGAAGGAAGTGCCGCGTAA 2802
clone2WT CAACTATGGTTATTTCGGAAAGGCAGCCTCTCTGAGTGAACTGAAGGAAGTGCCGCGTAA 2802
clone1wt AAACATCACCCTCCTCAGGGCGTTGGGACACGGTGCCTTTGGGGAAGTGTATGAAGGACA 2940
clone3wt AAACATCACCCTCCTCAGGGCGTTGGGACACGGTGCCTTTGGGGAAGTGTATGAAGGACA 2862
clone2WT AAACATCACCCTCCTCAGGGCGTTGGGACACGGTGCCTTTGGGGAAGTGTATGAAGGACA 2862
SP6-4b


clone1wt GGTTTTGGGTATGAATGGAGAAAACACAGCCATGCAAGTCGCTATAAAGACTCTTCCAGA 3000
clone3wt GGTTTTGGGTATGAATGGAGAAAACACAGCCATGCAAGTCGCTATAAAGACTCTTCCAGA 2922
clone2WT GGTTTTGGGTATGAATGGAGAAAACACAGCCATGCAAGTCGCTATAAAGACTCTTCCAGA 2922
clone1wt AATCTGCTCTGAGCAGGATGAGATGGATTTCCTAATGGAGGCTCTGATCATGAGTAAGTT 3060
clone3wt AATCTGCTCTGAGCAGGACGAGATGGATTTCCTAATGGAGGCTCTGATCATGAGTAAGTT 2982
clone2WT AATCTGCTCTGAGCAGGATGAGATGGATTTCCTAATGGAGGCTCTGATCATGAGTAAGTT 2982
clone1wt CAGCCATCAGAACATTGTCCGCTGCATTGGCGTCAGTCTGCAGATCCTGCCACGCTTCAT 3120
clone3wt CAGCCATCAGAACATTGTCCGCTGCATTGGCGTCAGTCTGCAGATCCTGCCACGCTTCAT 3042
clone2WT CAGCCATCAGAACATTGTCCGCTGCATTGGCGTCAGTCTGCAGATCCTGCCACGCTTCAT 3042
clone1wt CCTGCTGGAGCTCATGACAGGAGGAGACATGAAGAGCTTCTTGAGACTCAATCGACCCAG 3180
clone3wt CCTGCTGGAGCTCATGACAGGAGGAGACATGAAGAGCTTCTTGAGACTCAATCGACCCAG 3102
clone2WT CCTGCTGGAGCTCATGACAGGAGGAGACATGAAGAGCTTCTTGAGACTCAATCGACCCAG 3102
clone1wt AACTAACCATTCGTCTTCTCTAAGCATGCTGGAGCTTCTTCATATGGCCAGAGACATCGC 3240
clone3wt AACTAACCATTCGTCTTCTCTAAGCATGCTGGAGCTTCTTCATATGGCCAGAGACATCGC 3162
clone2WT AACTAACCATTCGTCTTCTCTAAGCATGCTGGAGCTTCTTCATATGGCCAGAGACATCGC 3162
clone1wt TCTTGGCTGTCGCTACCTTGAAGAAAACCACTTCATCCACAGGGACATCGCCGCTCGCAA 3300
clone3wt TCTTGGCTGTCGCTACCTTGAAGAAAACCACTTCATCCACAGGGACATCGCCGCTCGCAA 3222
clone2WT TCTTGGCTGTCGCTACCTTGAAGAAAACCACTTCATCCACAGGGACATCGCCGCTCGCAA 3222
SP6-4RC
clone1wt CTGCCTTTTGACTTGTCCTGGTCCAGACAGAGTGGCTAAAATTGGAGATTTCGGGATGGC 3360
clone3wt CTGCCTTTTGACTTGTCCTGGTCCAGACAGAGTGGCTAAAATTGGAGATTTCGGGATGGC 3282
clone2WT CTGCCTTTTGACTTGTCCTGGTCCAGACAGAGTGGCTAAAATTGGAGATTTCGGGATGGC 3282
SP6-4
clone1wt CCGAGATATTTACAGGGCCAGTTACTATAGGAAGGGTGGCCGTGCCATGCTGCCAGTCAA 3420
clone3wt CCGAGATATTTACAGGGCCAGTTACTATAGGAAGGGTGGCCGTGCCATGCTGCCAGTCAA 3342
clone2WT CCGAGATATTTACAGGGCCAGTTACTATAGGAAGGGTGGCCGTGCCATGCTGCCAGTCAA 3342
clone1wt ATGGATGCCACCTGAAGCTTTCCTAGAGGGCATTTTTACATGCAAGACTGACACCTGGTC 3480
clone3wt ATGGATGCCACCTGAAGCTTTCCTAGAGGGCATTTTTACATGCAAGACTGACACCTGGTC 3402
clone2WT ATGGATGCCACCTGAAGCTTTCCTAGAGGGCATTTTTACATGCAAGACTGACACCTGGTC 3402
SP6-3RC
clone1wt ATTCGGGGTACTGCTGTGGGAGATTTTCTCTCTTGGGTACATGCCTTATCCCTGCAAAAC 3540
clone3wt ATTCGGGGTACTGCTGTGGGAGATTTTCTCTCTTGGGTACATGCCTTATCCCTGCAAAAC 3462
clone2WT ATTCGGGGTACTGCTGTGGGAGATTTTCTCTCTTGGGTACATGCCTTATCCCTGCAAAAC 3462
SP6-3
clone1wt TAACCAGGAAGTGCTGGAGTTTGTAACTGGTGGAGGTCGCATGGATCCACCTAAGAGCTG 3600
clone3wt TAACCAGGAAGTGCTGGAGTTTGTAACTGGTGGAGGTCGCATGGATCCACCTAAGAGCTG 3522
clone2WT TAACCAGGAAGTGCTGGAGTTTGTAACTGGTGGAGGTCGCATGGATCCACCTAAGAGCTG 3522
clone1wt CCCTGGGCCTGTTTATCGGATCATGACACAGTGTTGGCAACACTGTCCAGAACACAGACC 3660
clone3wt CCCTGGGCCTGTTTATCGGATCATGACACAGTGTTGGCAACACTGTCCAGAACACAGACC 3582
clone2WT CCCTGGGCCTGTTTATCGGATCATGACACAGTGTTGGCAACACTGTCCAGAACACAGACC 3582

```
clone1wt        AAACTTCACTACTATTTTAGAGAGGATCAACTACTGTACACAGGATCCTGATGTCATCAA 3720
clone3wt        AAACTTCACTACTATTTTAGAGAGGATCAACTACTGTACACAGGATCCTGATGTCATCAA 3642
clone2WT        AAACTTCACTACTATTTTAGAGAGGATCAACTACTGTACACAGGATCCTGATGTCATCAA 3642
                ************************************************************

clone1wt        CACCCCTCTTCCAGTGGAGTGTGGTCCTCCTGTTGAAGAAGAAGGGGGCACTGTGATCCG 3780
clone3wt        CACCCCTCTTCCAGTGGAGTGTGGTCCTCCTGTTGAAGAAGAAGGGGGCACTGTGATCCG 3702
clone2WT        CACCCCTCTTCCAGTGGAGTGTGGTCCTCCTGTTGAAGAAGAAGGGGGCACTGTGATCCG 3702
                ************************************************************

clone1wt        CCCTGACGGATCAGGCAGCATGACCCCTCTTCTAGTTGCCCGCTCTTTGTCCCAGGATGC 3840
clone3wt        CCCTGACGGATCAGGCAGCATGACCCCTCTTCTAGTTGCCCGCTCTTTGTCCCAGGATGC 3762
clone2WT        CCCTGACGGATCAGGCAGCATGACCCCTCTTCTAGTTGCCCGCTCTTTGTCCCAGGATGC 3762
                ************************************************************

clone1wt        TTCCCCTCGAGCAAGCATCACCAGTGTCACTCCACAGGCCCTCAAACCCCGTTTGCAGCT 3900
clone3wt        TTCCCCTCGAGCAAGCATCACCAGTGTCACTCCACAGGCCCTCAAACCCCGTTTGCAGCT 3822
clone2WT        TTCCCCTCGAGCAAGCATCACCAGTGTCACTCCACAGGCCCTCAAACCCCGTTTGCAGCT 3822
                ************************************************************

clone1wt        ACAAAGACCAGTCCACCTCACACAAGAAGTGGGCACATACCGAGAGACACTGGAGCCCTG 3960
clone3wt        ACAAAGACCAGTCCACCTCACACAAGAAGTGGGCACATACCGAGAGACACTGGAGCCCTG 3882
clone2WT        ACAAAGACCAGTCCACCTCACACAAGAAGTGGGCACATACCGAGAGACACTGGAGCCCTG 3882
                ************************************ SP6-3b  <-- ********

                                                          -->

clone1wt        TTGGGCAGAGCCTGTTCCTGCTTCAGGAGTCTGTCCAGGACCATGGCTTCAAGTCCCAGA 4020
clone3wt        TTGGGCAGAGCCTGTTCCTGCTTCAGGAGTCTGTCCAGGACCATGGCTTCAAGTCCCAGA 3942
clone2WT        TTGGGCAGAGCCTGTTCCTGCTTCAGGAGTCTGTCCAGGACCATGGCTTCAAGTCCCAGA 3942
                ************************************************** SP6-2

SP6-2RC
clone1wt        GCATCGGCCATGCTCCAGAAGTAGTTCATCATCGGGCAGCCAGAAGCTAAAGAACAAGAC 4080
clone3wt        GCATCGGCCATGCTCCAGAAGTAGTTCATCATCGGGCAGCCAGAAGCTAAAGAACAAGAC 4002
clone2WT        GCATCGGCCATGCTCCAGAAGTAGTTCATCATCGGGCAGCCAGAAGCTAAAGAACAAGAC 4002
                <-- ************************************************************

clone1wt        TAAAAACCTCTGGAACCCCACCTACGGTTCCTGGGTCCTAGAGAGCTTTGGACGTGGAAA 4140
clone3wt        TAAAAACCTCTGGAACCCCACCTACGGTTCCTGGGTCCTAGAGAGCTTTGGACGTGGAAA 4062
clone2WT        TAAAAACCTCTGGAACCCCACCTACGGTTCCTGGGTCCTAGAGAGCTTTGGACGTGGAAA 4062
                ************************************************************

clone1wt        GTCAGCTCTGTGCCACACTCAGTCTATGCCTCTCTCCTGCAACCCCACATCTGTTTCTGC 4200
clone3wt        GTCAGCTCTGTGCCACACTCAGTCTATGCCTCTCTCCTGCAACCCCACATCTGTTTCTGC 4122
clone2WT        GTCAGCTCTGTGCCACACTCAGTCTATGCCTCTCTCCTGCAACCCCACATCTGTTTCTGC 4122
                ************************************************************

clone1wt        CCCGTCCTCTACCTCGGAGCACACAGACCCTGTAGTAGAAGTTAATGCAAATGTTTCAGC 4260
clone3wt        CCCGTCCTCTACCTCGGAGCACACAGACCCTGTAGTAGAAGTTAATGCAAATGTTTCAGC 4182
clone2WT        CCCGTCCTCTACCTCGGAGCACACAGACCCTGTAGTAGAAGTTAATGCAAATGTTTCAGC 4182
                ************************************************************

                --> N3GSP
clone1wt        TTCCCCACCTCCATCTGCTGCACCCTCACAAACTACGTTGACTCCCACAGCGGCTCCCTC 4320
clone3wt        TTCCCCACCTCCATCTGCTGCACCCTCACAAACTACGTTGACTCCCACAGCGGCTCCCTC 4242
clone2WT        TTCCCCACCTCCATCTGCTGCACCCTCACAAACTACGTTGACTCCCACAGCGGCTCCCTC 4242
                ************************************************************

clone1wt        TAGGAAGAGTCCCACAGGGGCTGCTGGTGTCTCACTTGCCACTGTTATGGACTTAGCAAA 4380
clone3wt        TAGGAAGAGTCCCACAGGGGCTGCTGGTGTCTCACTTGCCACTGTTATGGACTTAGCAAA 4302
clone2WT        TAGGAAGAGTCCCACAGGGGCTGCTGGTGTCTCACTTGCCACTGTTATGGACTTAGCAAA 4302
                ************************************************************

clone1wt        GCTTCAGAGCTTCCCATGTGGCAACGTAAACTATGCGTATGATGAACAGAGCTACGAGAC 4440
clone3wt        GCTTCAGAGCTTCCCATGTGGCAACGTAAACTATGCGTATGATGAACAGAGCTACGAGAC 4362
clone2WT        GCTTCAGAGCTTCCCATGTGGCAACGTAAACTATGCGTATGATGAACAGAGCTACGAGAC 4362
                ************************************************************
```


clone1wt TGAGAGTTTGCCTGTGGTTGTGTCCAAATCTCTGGAGCCCAGCACCAGCTCTGCAGCTAC 4500
clone3wt TGAGAGTTTGCCTGTGGTTGTGTCCAAATCTCTGGAGCCCAGCACCAGCTCTGCAGCTAC 4422
clone2WT TGAGAGTTTGCCTGTGGTTGTGTCCAAATCTCTGGAGCCCAGCACCAGCTCTGCAGCTAC 4422
clone1wt ATCTTCACTGGTTGCCCTCAGCCAAGCCAGCAGCTTTACCCACAAACCATTGGTGAAGCG 4560
clone3wt ATCTTCACTGGTTGCCCTCAGCCAAGCCAGCAGCTTTACCCACAAACCATTGGTGAAGCG 4482
clone2WT ATCTTCACTGGTTGCCCTCAGCCAAGCCAGCAGCTTTACCCACAAACCATTGGTGAAGCG 4482
clone1wt TCATGCCAGTTATGGACATGAGGATGTAAGGAGGTACACCCAGCCTGAGAAGCCCACCAG 4620
clone3wt TCATGCCAGTTATGGACATGAGGATGTAAGGAGGTACACCCAGCCTGAGAAGCCCACCAG 4542
clone2WT TCATGCCAGTTATGGACATGAGGATGTAAGGAGGTACACCCAGCCTGAGAAGCCCACCAG 4542
SP6-2b
SP6-1
SP6-1RC
STOP
clone1wt GGACAGGGATTCTGGCTTTTCCTTGTCTGAAGACCTGAGTGTCACCCCTGTGTAGCAATG 4680
clone3wt GGACAGGGATTCTGGCTTTTCCTTGTCTGAAGACCTGAGTGTCACCCCTGTGTAGCAATG 4602
clone2WT GGACAGGGATTCTGGCTTTTCCTTGTCTGAAGACCTGAGTGTCACCCCTGTGTAGCAATG 4602
clone1wt AGGAGTGGTCCGACTTAGAAAATGCCTATTGTTTTTGGTGTTGAATAGTTTTGAGGTGTT 4740
clone3wt AGGAGTGGTCCGACTTAGAAAATGCCTATTGTTTTTGGTGTTGAATAGTTTTGAGGTGTT 4662
clone2WT AGGAGTGGTCCGACTTAGAAAATGCCTATTGTTTTTGGTGTTGAATAGTTTTGAGGTGTT 4662
clone1wt TTTTTTATGTGAACTGCAATGGATTAGCTGGAGCTGGGGCGTCAAATCCTGATCCAGCTT 4800
clone3wt TTTTTTATGTGAACTGCAATGGATTAGCTGGAGCTGGGGCGTCAAATCCTGATCCAGCTT 4722
clone2WT TTTTTTATGTGAACTGCAATGGATTAGCTGGAGCTGGGGCGTCAAATCCTGATCCAGCTT 4722
clone1wt CAACACACTTGCCTGGAGTTTTCAATAATTAGTCTGAAGACCTTGATTTGAAAACTTTGC 4860
clone3wt CAACACACTTGCCTGGAGTTTTCAATAATTAGTCTGAAGACCTTGATTTGAAAACTTTGC 4782
clone2WT CAACACACTTGCCTGGAGTTTTCAATAATTAGTCTGAAGACCTTGATTTGAAAACTTTGC 4782
clone1wt TGGATTAAACAGCTGGTTTTGGACACCCCTGAGCTACAGGCAGAGTGGCAGGTTAAGGCT 4920
clone3wt TGGATTAAACAGCTGGTTTTGGACACCCCTGAGCTACAGGCAGAGTGGCAGGTTAAGGCT 4842
clone2WT TGGATTAAACAGCTGGTTTTGGACACCCCTGAGCTACAGGCAGAGTGGCAGGTTAAGGCT 4842
clone1wt GGAATATACTACACTATAGTGGTCTTCAGTGATTCGACCCAAGCGTGTTCCTTGGACATG 4980
clone3wt GGAATATACTACACTATAGTGGTCTTCAGTGATTCGACCCAAGCGTGTTCCTTGGACATG 4902
clone2WT GGAATATACTACACTATAGTGGTCTTCAGTGATTCGACCCAAGCGTGTTCCTTGGACATG 4902
3'LD1
clone1wt GGGCTTGGGTTACTGAATCTAAATGTGCTTTTTGCAAATGGACCTGATAACACCCAAAAC 5040
clone3wt GGGCTTGGGTTACTGAATCTAAATGTGCTTTTTGCAAATTGACCTGATAACACCC----- 4957
clone2WT GGGCTTGGGTTACTGAATCTAAATGTGCTTTTTGCAAATGGACCTGATAACACCCAAAAC 4962
clone1wt GTTAATGAGATGACA-TGATTTTGCCCATAAAATTCCAGTAAACACCAAACCACGTAACT 5099
clone3wt ----------
clone2WT GTTAATGAGATGACA-TGATTTTGCCCATAAAATTCCAGTAAACACCAAACCACGTAACT 5021
clone1wt TCACTTTCAAAATCACACTGTACCAGATCATTCCATACATAAATCCCCCAGACACATGCA 5159
clone3wt ----------
clone2WT TCACTTTCAAAATCACACTGTACCAGATCATTCCATACATAAATCCCCCAGACACATGCA 5081
SP6-1b
clone1wt CCCCATTGATAGAAGACAAATGACAAATGGAGTCAAGAGTCTGATCAACACGTTTATTGT 5219
clone3wt ----------
clone2WT CCCCATTGATAGAAGACAAATGACAAATGGAGTCAAGAGTCTGATCAACACGTTTATTGT 5141


clone1wt TTTCTAACTGAGTGTATCCATAATCTTAAGCTTTAAAGTCATTGTATTCATTAGATATTA 5279
clone3wt ----------
clone2WT TTTCTAACTGAGTGTATCCATAATCTTAAGCTTTAAAGTCATTGTATTCATTAGATATTA 5201
clone1wt AACACTACACATTAAACACTACACAAACTTTCATGCAACGTTTTGGATCTTGTGCAAATC 5339
clone3wt ----------
clone2WT AACACTACACATTAAACACTACACAAACTTTCATGCAACGTTTTGGATCTTGTGCAAATC 5261
clone1wt AGCAGACATTTGGTAGCTAATGTTGACTCAAGAGTGAAAAATTGGTATACAAATCTGGTT 5399
clone3wt ----------
clone2WT AGCAGACATTTGGTAGCTAATGTTGACTCAAGAGTGAAAAATTGGTATACAAATCTGGTT 5321
clone1wt AAATATTGTGTAATATATTCCAGCCTTGAAGTACTGTCAGTTGATATCCAAGTGCGTGAA 5459
clone3wt ----------
clone2WT AAATATTGTGTAATATATTCCAGCCTTGAAGTACTGTCAGTTGATATCCAAGTGCGTGAA 5381
3'LD2
clone1wt ATACCAAAACACCAAAGACATGAGGATACATGATTAATTTGCCATTAAGAGACACCTCTC 5519
clone3wt ----------
clone2WT ATACCA---------- 5387
clone1wt CTGAAAATGGTTGATGGTTTCTATGAAAACTTCACCTTGGCTTATGGCAGCCCTTTTACA 5579
clone3wt ----------
clone2WT ----------
3'LD3
clone1wt CTGGATTCACGTCTACTGTCAACACTGTGAAAAC 5613
clone3wt ----------
clone2WT ----------

2. CLUSTAL W (1.82) multiple sequence alignment between translated clones 1 to 3
Different domains are annotated in the same colours as diagram in Figure 41

```
clone1          MDYITRQTFVKLALFIFTVVRSSCALLEKAAEDPVHPNPLQSSPAEDSDVSFCDFESPCS 60
clone3          MDYITRQTFVKLALFIFTVVRSSCALLEKAAEDPVHPNPLQSSPAEDSDVSFCDFESPCS 60
clone2          MDYITRQTFVKLALFIFTVVRSSCALLEKAAEDPVHPNPLQSSPAEDSDVSFCDFESPCS 60
                ************************************************************

                                                                  ---------> MAM1

clone1          WTLSSHSTGGDWFITSAQQHRSNRRDTQPIRDYSTGKSEGHFLLLKPSSSHLSAGRCSFH 120
clone3          WTLSSHSTGGDWFITSAQQHRSNRRDTQPIRDYSTGKSEGHFLLLKPSSSHLSAGRCSFH 120
clone2          WTLSSHSTGGDWFITSAQQHRSNRRDTQPIRDYSTGKSEGHFLLLKPSSSHLSAGRCSFH 120
                ************************************************************

clone1          MTSPVVLSSGPFCHLQLARFQPEPHAGNISAFVKHTDSIDIKPIDLTIKEQESDSSQWEV 180
clone3          MTSPVVLSSGPFCHLQLARFQPEPHAGNISAFVKHTDSIDIKPIDLTIKEQESDSSQWEV 180
clone2          MTSPVVLSSGPFCHLQLARFQPEPHAGNISAFVKHTDSIDIKPIDLTIKEQESDSSQWEV 180
                ************************************************************

clone1          LEAVIGQLNEPFQVTVQYSACSSHEVGFLAFDSLELKNCVMGDDYVDLGSDCEKYSSLQC 240
clone3          LEAVIGQLNEPFQVTVQYSACSSHEVGFLAFDSLELKNCVMGDDYVDLGSDCEKYSSLQC 240
clone2          LEAVIGQLNEPFQVTVQYSACSSHEVGFLAFDSLELKNCVMGDDYVDLGSDCEKYSSLQC 240
                ************************************************************

clone1          HSGGCIEKQRVCDFHTDCPEGEDEGLICSTLPLGSYCSFELGSCGWSAADTQSSWRLVSG 300
clone3          HSGGCIEKQRVCDFHTDCPEGEDEGLICSTLPLGSYCSFELGSCGWSAADTQSSWRLVSG 300
clone2          HSGGCIEKQRVCDFHTDCPEGEDEGLICSTLPLGSYCSFELGSCGWSAADTQSSWRLVSG 300
                ****************************************************************

                ---------> LDLa                                  ---------> MAM2

clone1          QQLIEDTHLLGTTLKNTQGHFLFLKVRGHGDEREALVQSPALPSTISNQDCQLQFSLYRY 360
clone3          QQLIEDTHLLGTTLKNTQGHFLFLKVRGHGDEREALVQSPALPSTISNQDCQLQFSLYRY 360
clone2          QQLIEDTHLLGTTLKNTQGHFLFLKVRGHGDEREALVQSPALPSTISNQDCQLQFSLYRY 360
                ************************************************************

clone1          GDFNGTVLLSVVESGASAPALIWERSGHWKDAWQEITLPITEILNGFHLKVQAFWTSGSK 420
clone3          GDFNGTVLLSVVESGASAPALIWERSGHWKDAWQEITLPITEILNGFHLKVQAFWTSGSK 420
clone2          GDFNGTVLLSVVESGASAPALIWERSGHWKDAWQEITLPITEILNGFHLKVQAFWTSGSK 420
                ************************************************************

clone1          ADIALDDISLSAACFDTELNELLHEGLPHDLDFSPLPEPSASEASPITWWFTSCGASGPF 480
clone3          ADIALDDISLSAACFDTELNELLHEGLPHDLDFSPLPEPSASEASPITWWFTSCGASGPF 480
clone2          ADIALDDISLSAACFDTELNELLHEGLPHDLDFSPLPEPSASEASPITWWFTSCGASGPF 480
                ************************************************************

clone1          GPTQAQCDSAYRNTNVSVVVGKEGPLRGVQMWKVPATNTYKISAYGAAGGKGAKNHNKRS 540
clone3          GPTQAQCDSAYRNTNVSVVVGKEGPLRGVQMWKVPATNTYKISAYGAAGGKGAKNHNKRS 540
clone2          GPTQAQCDSAYRNTNVSVVVGKEGPLRGVQMWKVPATNTYKISAYGAAGGKGAKNHNKRS 540
                ************************************************************

clone1          HGVFISATFPLEKGDILYILIGHQGEDACPGRNPQTHKICLGESSVIEDGFDSDGSALKW 600
clone3          HGVFISATFPLEKGDILYILIGHQGEDACPGRNPQTHKICLGESSVIEDGFDSDGSALKW 600
clone2          HGVFISATFPLEKGDILYILIGHQGEDACPGRNPQTHKICLGESSVIEDGFDSDGSALKW 600
                ************************************************************

clone1          AGGGGGGGGATYIYRMENGQPLPLLIAAGGGGKAYLEDPESSQDQSFREQYENDTTVSGV 660
clone3          AGGGGGGGGATYIYRMENGQPLPLLIAAGGGGKAYLEDPESSQDQSFREQYENDTTVSGV 660
clone2          AGGGGGGGGATYIYRMENGQPLPLLIAAGGGGKAYLEDPESSQDQSFREQYENDTTVSGV 660
                ************************************************************

                ---------> G-rich

clone1          SGRSGAAGGGGGWSDVSSLSWAGKSLVEGGQGGSSCPEALSVLGWATFGGFGGGGGACSA 720
clone3          SGRSGAAGGGGGWSDVSSLSWAGKSLVEGGQGGSSCPEALSVLGWATFGGFGGGGGACSA 720
clone2          SGRSGAAGGGGGWSDVSSLSWAGKSLVEGGQGGSSCPEALSVLGWATFGGFGGGRGACSA 720
                ********************************************************* *****
```


clone1 GGGGGGYRGGDAPLLDDISADGQDGLSFVHPMGKIFLQPLAAMESHGEAEIVVYLNCSHC 780
clone3 GGGGGGYRGGDAPLLDDISADGQDGLSFVHPMGKIFLQPLAAMESHGEAEIVVYLNCSHC 780
clone2 GGGGGGYRGGDAPLLDDISADGQDGLSFVHPMGKIFLQPLAAMESHGEAEIVVYLNCSHC 780
clone1 KTQSCKRDEDTKLILCLCDSDEVLAPDNVTCAGTKHSLCQFINKHLQHNSSPLVCPPLVV 840
clone3 KTQSCKRDEDTKLILCLCDSDEVLAPDNVTCA----------VV 814
clone2 KTQSCKRDEDTKLILCLCDSDEVLAPDNVTCA----------VV 814
clone1 PMGSLADGPPSLVFIMAVIVSTVVTGVVLTCASLTLIYYRKKNHLHAVRIRLQSPEYKLS 900
clone3 PMGSLADGPPSLVFIMAGIVSTVVTGVVLTCASLTLIYYRKKNHLHAVRIRLQSPEYKLS 874
clone2 PMGSLADGPPSLVFIMAGIVSTVVTGVVLTCASLTLIYYRKKNHLHAVRIRLQSPEYKLS 874
TM
PTK
clone1 KIRSSTIMTDYNPNYGYFGKAASLSELKEVPRKNITLLRALGHGAFGEVYEGQVLGMNGE 960
clone3 KIRSSTIMTDYNPNYGYFGKAASLSELKEVPRKNITLLRALGHGAFGEVYEGQVLGMNGE 934
clone2 KIRSSTIMTDYNPNYGYFGKAASLSELKEVPRKNITLLRALGHGAFGEVYEGQVLGMNGE 934
ATP
clone1 NTAMQVAIKTLPEICSEQDEMDFLMEALIMSKFSHQNIVRCIGVSLQILPRFILLELMTG 1020
clone3 NTAMQVAIKTLPEICSEQDEMDFLMEALIMSKFSHQNIVRCIGVSLQILPRFILLELMTG 994
clone2 NTAMQVAIKTLPEICSEQDEMDFLMEALIMSKFSHQNIVRCIGVSLQILPRFILLELMTG 994
clone1 GDMKSFLRLNRPRTNHSSSLSMLELLHMARDIALGCRYLEENHFIHRDIAARNCLLTCPG 1080
clone3 GDMKSFLRLNRPRTNHSSSLSMLELLHMARDIALGCRYLEENHFIHRDIAARNCLLTCPG 1054
clone2 GDMKSFLRLNRPRTNHSSSLSMLELLHMARDIALGCRYLEENHFIHRDIAARNCLLTCPG 1054
clone1 PDRVAKIGDFGMARDIYRASYYRKGGRAMLPVKWMPPEAFLEGIFTCKTDTWSFGVLLWE 1140
clone3 PDRVAKIGDFGMARDIYRASYYRKGGRAMLPVKWMPPEAFLEGIFTCKTDTWSFGVLLWE 1114
clone2 PDRVAKIGDFGMARDIYRASYYRKGGRAMLPVKWMPPEAFLEGIFTCKTDTWSFGVLLWE 1114
clone1 IFSLGYMPYPCKTNQEVLEFVTGGGRMDPPKSCPGPVYRIMTQCWQHCPEHRPNFTTILE 1200
clone3 IFSLGYMPYPCKTNQEVLEFVTGGGRMDPPKSCPGPVYRIMTQCWQHCPEHRPNFTTILE 1174
clone2 IFSLGYMPYPCKTNQEVLEFVTGGGRMDPPKSCPGPVYRIMTQCWQHCPEHRPNFTTILE 1174
clone1 RINYCTQDPDVINTPLPVECGPPVEEEGGTVIRPDGSGSMTPLLVARSLSQDASPRASIT 1260
clone3 RINYCTQDPDVINTPLPVECGPPVEEEGGTVIRPDGSGSMTPLLVARSLSQDASPRASIT 1234
clone2 RINYCTQDPDVINTPLPVECGPPVEEEGGTVIRPDGSGSMTPLLVARSLSQDASPRASIT 1234
clone1 SVTPQALKPRLQLQRPVHLTQEVGTYRETLEPCWAEPVPASGVCPGPWLQVPEHRPCSRS 1320
clone3 SVTPQALKPRLQLQRPVHLTQEVGTYRETLEPCWAEPVPASGVCPGPWLQVPEHRPCSRS 1294
clone2 SVTPQALKPRLQLQRPVHLTQEVGTYRETLEPCWAEPVPASGVCPGPWLQVPEHRPCSRS 1294
clone1 SSSSGSQKLKNKTKNLWNPTYGSWVLESFGRGKSALCHTQSMPLSCNPTSVSAPSSTSEH 1380
clone3 SSSSGSQKLKNKTKNLWNPTYGSWVLESFGRGKSALCHTQSMPLSCNPTSVSAPSSTSEH 1354
clone2 SSSSGSQKLKNKTKNLWNPTYGSWVLESFGRGKSALCHTQSMPLSCNPTSVSAPSSTSEH 1354
clone1 TDPVVEVNANVSASPPPSAAPSQTTLTPTAAPSRKSPTGAAGVSLATVMDLAKLQSFPCG 1440
clone3 TDPVVEVNANVSASPPPSAAPSQTTLTPTAAPSRKSPTGAAGVSLATVMDLAKLQSFPCG 1414
clone2 TDPVVEVNANVSASPPPSAAPSQTTLTPTAAPSRKSPTGAAGVSLATVMDLAKLQSFPCG 1414
clone1 NVNYAYDEQSYETESLPVVVSKSLEPSTSSAATSSLVALSQASSFTHKPLVKRHASYGHE 1500
clone3 NVNYAYDEQSYETESLPVVVSKSLEPSTSSAATSSLVALSQASSFTHKPLVKRHASYGHE 1474
clone2 NVNYAYDEQSYETESLPVVVSKSLEPSTSSAATSSLVALSQASSFTHKPLVKRHASYGHE 1474

```
clone1          DVRRYTQPEKPTRDRDSGFSLSEDLSVTPV 1530
clone3          DVRRYTQPEKPTRDRDSGFSLSEDLSVTPV 1504
clone2          DVRRYTQPEKPTRDRDSGFSLSEDLSVTPV 1504
                ******************************
```

3. Coding sequence with correspondent aminoacid translation. The 1200bp fragment used for the riboprobe synthesis is annotated in red.

```
1     ATG GAT TAT ATT ACT CGA CAA ACT TTT GTA AAA CTT GCC CTT TTC   45
1      M   D   Y   I   T   R   Q   T   F   V   K   L   A   L   F    15

46    ATC TTC ACA GTT GTG AGG AGC AGT TGT GCT CTG TTG GAA AAA GCT   90
16     I   F   T   V   V   R   S   S   C   A   L   L   E   K   A    30

91    GCA GAA GAC CCA GTG CAT CCA AAT CCT CTT CAA TCC AGT CCA GCA   135
31     A   E   D   P   V   H   P   N   P   L   Q   S   S   P   A    45

136   GAA GAC AGC GAT GTT TCG TTC TGT GAT TTC GAG TCT CCG TGT TCT   180
46     E   D   S   D   V   S   F   C   D   F   E   S   P   C   S    60

181   TGG ACA TTA TCC AGT CAC AGC ACT GGA GGA GAC TGG TTT ATC ACT   225
61     W   T   L   S   S   H   S   T   G   G   D   W   F   I   T    75

226   TCA GCA CAG CAG CAC AGA TCT AAC AGA CGC GAC ACC CAG CCT ATC   270
76     S   A   Q   Q   H   R   S   N   R   R   D   T   Q   P   I    90

271   AGA GAT TAC TCC ACT GGA AAA TCT GAA GGA CAT TTT CTG CTG CTG   315
91     R   D   Y   S   T   G   K   S   E   G   H   F   L   L   L   105

316   AAG CCC AGC TCC AGT CAT CTG TCT GCA GGC AGA TGT AGT TTC CAC   360
106    K   P   S   S   S   H   L   S   A   G   R   C   S   F   H   120

361   ATG ACC AGT CCT GTA GTT CTC AGC AGT GGG CCA TTT TGT CAT CTC   405
121    M   T   S   P   V   V   L   S   S   G   P   F   C   H   L   135

406   CAA CTC GCT CGC TTT CAA CCA GAG CCA CAC GCT GGA AAC ATA TCA   450
136    Q   L   A   R   F   Q   P   E   P   H   A   G   N   I   S   150

451   GCC TTC GTG AAG CAC ACC GAC TCT ATC GAC ATC AAA CCC ATT GAC   495
151    A   F   V   K   H   T   D   S   I   D   I   K   P   I   D   165

496   CTT ACA ATC AAA GAA CAA GAA AGC GAC AGC TCG CAG TGG GAG GTC   540
166    L   T   I   K   E   Q   E   S   D   S   S   Q   W   E   V   180

541   TTG GAG GCT GTG ATT GGT CAG TTG AAT GAG CCC TTC CAG GTG ACA   585
181    L   E   A   V   I   G   Q   L   N   E   P   F   Q   V   T   195

586   GTG CAG TAC TCT GCC TGT AGC AGT CAT GAA GTT GGG TTT CTG GCA   630
196    V   Q   Y   S   A   C   S   S   H   E   V   G   F   L   A   210

631   TTC GAT TCA CTG GAG CTG AAG AAC TGT GTC ATG GGT GAT GAT TAT   675
211    F   D   S   L   E   L   K   N   C   V   M   G   D   D   Y   225

676   GTG GAT TTG GGT TCC GAC TGT GAG AAA TAC TCA TCT CTT CAA TGT   720
226    V   D   L   G   S   D   C   E   K   Y   S   S   L   Q   C   240

721   CAT TCT GGA GGC TGC ATT GAG AAA CAA AGA GTG TGT GAT TTC CAC   765
241    H   S   G   G   C   I   E   K   Q   R   V   C   D   F   H   255

766   ACT GAC TGC CCA GAA GGG GAG GAC GAG GGC TTA ATA TGC AGC ACT   810
256    T   D   C   P   E   G   E   D   E   G   L   I   C   S   T   270

811   TTA CCT TTG GGT TCA TAC TGT TCA TTT GAG TTG GGC TCC TGT GGT   855
271    L   P   L   G   S   Y   C   S   F   E   L   G   S   C   G   285

856   TGG TCA GCG GCT GAC ACA CAG TCT TCT TGG AGA CTA GTT AGT GGA   900
286    W   S   A   A   D   T   Q   S   S   W   R   L   V   S   G   300

901   CAA CAG CTG ATT GAA GAC ACA CAC CTG CTG GGC ACA ACT CTG AAG   945
301    Q   Q   L   I   E   D   T   H   L   L   G   T   T   L   K   315

946   AAC ACT CAA GGG CAT TTC CTT TTC TTG AAG GTC AGA GGT CAC GGT   990
316    N   T   Q   G   H   F   L   F   L   K   V   R   G   H   G   330
```

```
991   GAT GAA AGA GAG GCG CTG GTT CAA AGC CCT GCA TTG CCT TCA ACT   1035
331    D   E   R   E   A   L   V   Q   S   P   A   L   P   S   T   345

1036  ATA TCC AAT CAG GAC TGT CAG TTG CAG TTT TCT CTC TAC CGG TAT   1080
346    I   S   N   Q   D   C   Q   L   Q   F   S   L   Y   R   Y   360

1081  GGG GAT TTC AAT GGA ACA GTG CTT TTG TCT GTG GTG GAG AGC GGA   1125
361    G   D   F   N   G   T   V   L   L   S   V   V   E   S   G   375

1126  GCC TCA GCA CCA GCA CTG ATT TGG GAG AGA TCC GGA CAC TGG AAA   1170
376    A   S   A   P   A   L   I   W   E   R   S   G   H   W   K   390

1171  GAC GCC TGG CAA GAG ATC ACT CTG CCG ATC ACA GAG ATT TTA AAT   1215
391    D   A   W   Q   E   I   T   L   P   I   T   E   I   L   N   405

1216  GGC TTC CAT CTT AAA GTG CAG GCT TTC TGG ACT TCT GGC TCC AAG   1260
406    G   F   H   L   K   V   Q   A   F   W   T   S   G   S   K   420

1261  GCT GAC ATT GCA CTC GAT GAC ATT TCA TTA AGT GCA GCA TGT TTT   1305
421    A   D   I   A   L   D   D   I   S   L   S   A   A   C   F   435

1306  GAC ACA GAA CTG AAT GAA CTG CTT CAT GAA GGA CTA CCG CAT GAT   1350
436    D   T   E   L   N   E   L   L   H   E   G   L   P   H   D   450

1351  TTA GAC TTC AGT CCT CTT CCA GAA CCA TCA GCA TCA GAG GCG TCT   1395
451    L   D   F   S   P   L   P   E   P   S   A   S   E   A   S   465

1396  CCA ATT ACA TGG TGG TTT ACA TCC TGT GGA GCG AGT GGA CCT TTC   1440
466    P   I   T   W   W   F   T   S   C   G   A   S   G   P   F   480

1441  GGC CCG ACT CAA GCC CAG TGT GAT AGT GCC TAC CGA AAC ACC AAT   1485
481    G   P   T   Q   A   Q   C   D   S   A   Y   R   N   T   N   495

1486  GTC AGC GTT GTG GTG GGG AAA GAG GGT CCT CTC AGG GGT GTA CAG   1530
496    V   S   V   V   V   G   K   E   G   P   L   R   G   V   Q   510

1531  ATG TGG AAG GTC CCT GCT ACC AAC ACG TAC AAG ATT TCA GCC TAT   1575
511    M   W   K   V   P   A   T   N   T   Y   K   I   S   A   Y   525

1576  GGA GCT GCA GGA GGC AAA GGA GCA AAG AAC CAT AAC AAG CGC TCA   1620
526    G   A   A   G   G   K   G   A   K   N   H   N   K   R   S   540

1621  CAC GGT GTG TTC ATC TCA GCC ACC TTC CCC CTG GAG AAA GGT GAC   1665
541    H   G   V   F   I   S   A   T   F   P   L   E   K   G   D   555

1666  ATC CTC TAC ATC TTA ATT GGC CAC CAA GGA GAA GAT GCC TGT CCA   1710
556    I   L   Y   I   L   I   G   H   Q   G   E   D   A   C   P   570

1711  GGA AGG AAC CCC CAA ACC CAT AAA ATC TGT CTG GGC GAG TCG TCT   1755
571    G   R   N   P   Q   T   H   K   I   C   L   G   E   S   S   585

1756  GTG ATT GAA GAT GGT TTT GAC AGC GAC GGT TCA GCC TTG AAG TGG   1800
586    V   I   E   D   G   F   D   S   D   G   S   A   L   K   W   600

1801  GCA GGG GGT GGC GGC GGT GGA GGG GGA GCC ACA TAT ATA TAT CGG   1845
601    A   G   G   G   G   G   G   G   G   A   T   Y   I   Y   R   615

1846  ATG GAG AAC GGG CAA CCA CTT CCT TTG CTG ATT GCA GCT GGA GGG   1890
616    M   E   N   G   Q   P   L   P   L   L   I   A   A   G   G   630

1891  GGA GGA AAG GCC TAC CTT GAG GAT CCA GAA AGC AGC CAG GAC CAG   1935
631    G   G   K   A   Y   L   E   D   P   E   S   S   Q   D   Q   645

1936  AGT TTT CGG GAA CAG TAC GAG AAT GAC ACC ACC GTT TCT GGT GTC   1980
646    S   F   R   E   Q   Y   E   N   D   T   T   V   S   G   V   660
```

```
1981  AGT GGC AGA TCT GGA GCA GCA GGT GGA GGA GGT GGA TGG AGC GAC  2025
661    S   G   R   S   G   A   A   G   G   G   G   W   S   D    675

2026  GTG TCC TCT CTC TCA TGG GCA GGG AAA TCT CTT GTT GAA GGG GGT  2070
676    V   S   S   L   S   W   A   G   K   S   L   V   E   G   G    690

2071  CAA GGA GGC TCA TCC TGT CCT GAG GCT TTG TCA GTG CTG GGA TGG  2115
691    Q   G   G   S   S   C   P   E   A   L   S   V   L   G   W    705

2116  GCT ACA TTT GGA GGA TTC GGA GGA GGA GGA GGC GCG TGC TCC GCT  2160
706    A   T   F   G   G   F   G   G   G   G   G   A   C   S   A    720

2161  GGA GGG GGC GGT GGA GGA TAC AGA GGT GGT GAT GCA CCT CTT CTT  2205
721    G   G   G   G   G   G   Y   R   G   G   D   A   P   L   L    735

2206  GAC GAC ATC TCA GCA GAT GGA CAA GAT GGC CTC TCT TTT GTT CAC  2250
736    D   D   I   S   A   D   G   Q   D   G   L   S   F   V   H    750

2251  CCC ATG GGA AAG ATA TTT CTG CAA CCT CTG GCA GCC ATG GAG AGT  2295
751    P   M   G   K   I   F   L   Q   P   L   A   A   M   E   S    765

2296  CAC GGC GAA GCT GAA ATT GTG GTC TAT CTG AAC TGC AGC CAC TGT  2340
766    H   G   E   A   E   I   V   V   Y   L   N   C   S   H   C    780

2341  AAA ACC CAG AGC TGC AAG CGC GAT GAA GAC ACC AAG CTC ATC CTG  2385
781    K   T   Q   S   C   K   R   D   E   D   T   K   L   I   L    795

2386  TGC CTC TGC GAC AGC GAC GAG GTT CTG GCA CCA GAC AAC GTC ACC  2430
796    C   L   C   D   S   D   E   V   L   A   P   D   N   V   T    810

2431  TGC GCA GTG GTT CCT ATG GGC TCT CTG GCA GAC GGG CCT CCA TCC  2475
811    C   A   V   V   P   M   G   S   L   A   D   G   P   P   S    825

2476  CTG GTC TTC ATC ATG GCA GTG ATC GTG TCC ACA GTG GTG ACC GGT  2520
826    L   V   F   I   M   A   V   I   V   S   T   V   V   T   G    840

2521  GTC GTT CTG ACC TGT GCC AGC CTG ACT CTC ATA TAT TAT CGT AAG  2565
841    V   V   L   T   C   A   S   L   T   L   I   Y   Y   R   K    855

2566  AAG AAC CAC CTG CAT GCG GTC AGG ATT CGA CTG CAG AGT CCA GAG  2610
856    K   N   H   L   H   A   V   R   I   R   L   Q   S   P   E    870

2611  TAC AAG CTT AGC AAA ATT CGC TCG TCC ACC ATC ATG ACC GAC TAC  2655
871    Y   K   L   S   K   I   R   S   S   T   I   M   T   D   Y    885

2656  AAC CCC AAC TAT GGT TAT TTC GGA AAG GCA GCC TCT CTG AGT GAA  2700
886    N   P   N   Y   G   Y   F   G   K   A   A   S   L   S   E    900

2701  CTG AAG GAA GTG CCG CGT AAA AAC ATC ACC CTC CTC AGG GCG TTG  2745
901    L   K   E   V   P   R   K   N   I   T   L   L   R   A   L    915

2746  GGA CAC GGT GCC TTT GGG GAA GTG TAT GAA GGA CAG GTT TTG GGT  2790
916    G   H   G   A   F   G   E   V   Y   E   G   Q   V   L   G    930

2791  ATG AAT GGA GAA AAC ACA GCC ATG CAA GTC GCT ATA AAG ACT CTT  2835
931    M   N   G   E   N   T   A   M   Q   V   A   I   K   T   L    945

2836  CCA GAA ATC TGC TCT GAG CAG GAC GAG ATG GAT TTC CTA ATG GAG  2880
946    P   E   I   C   S   E   Q   D   E   M   D   F   L   M   E    960

2881  GCT CTG ATC ATG AGT AAG TTC AGC CAT CAG AAC ATT GTC CGC TGC  2925
961    A   L   I   M   S   K   F   S   H   Q   N   I   V   R   C    975

2926  ATT GGC GTC AGT CTG CAG ATC CTG CCA CGC TTC ATC CTG CTG GAG  2970
976    I   G   V   S   L   Q   I   L   P   R   F   I   L   L   E    990

2971  CTC ATG ACA GGA GGA GAC ATG AAG AGC TTC TTG AGA CTC AAT CGA  3015
991    L   M   T   G   G   D   M   K   S   F   L   R   L   N   R   1005
```

```
3016  CCC AGA ACT AAC CAT TCG TCT TCT CTA AGC ATG CTG GAG CTT CTT   3060
1006   P   R   T   N   H   S   S   S   L   S   M   L   E   L   L   1020

3061  CAT ATG GCC AGA GAC ATC GCT CTT GGC TGT CGC TAC CTT GAA GAA   3105
1021   H   M   A   R   D   I   A   L   G   C   R   Y   L   E   E   1035

3106  AAC CAC TTC ATC CAC AGG GAC ATC GCC GCT CGC AAC TGC CTT TTG   3150
1036   N   H   F   I   H   R   D   I   A   A   R   N   C   L   L   1050

3151  ACT TGT CCT GGT CCA GAC AGA GTG GCT AAA ATT GGA GAT TTC GGG   3195
1051   T   C   P   G   P   D   R   V   A   K   I   G   D   F   G   1065

3196  ATG GCC CGA GAT ATT TAC AGG GCC AGT TAC TAT AGG AAG GGT GGC   3240
1066   M   A   R   D   I   Y   R   A   S   Y   Y   R   K   G   G   1080

3241  CGT GCC ATG CTG CCA GTC AAA TGG ATG CCA CCT GAA GCT TTC CTA   3285
1081   R   A   M   L   P   V   K   W   M   P   P   E   A   F   L   1095

3286  GAG GGC ATT TTT ACA TGC AAG ACT GAC ACC TGG TCA TTC GGG GTA   3330
1096   E   G   I   F   T   C   K   T   D   T   W   S   F   G   V   1110

3331  CTG CTG TGG GAG ATT TTC TCT CTT GGG TAC ATG CCT TAT CCC TGC   3375
1111   L   L   W   E   I   F   S   L   G   Y   M   P   Y   P   C   1125

3376  AAA ACT AAC CAG GAA GTG CTG GAG TTT GTA ACT GGT GGA GGT CGC   3420
1126   K   T   N   Q   E   V   L   E   F   V   T   G   G   G   R   1140

3421  ATG GAT CCA CCT AAG AGC TGC CCT GGG CCT GTT TAT CGG ATC ATG   3465
1141   M   D   P   P   K   S   C   P   G   P   V   Y   R   I   M   1155

3466  ACA CAG TGT TGG CAA CAC TGT CCA GAA CAC AGA CCA AAC TTC ACT   3510
1156   T   Q   C   W   Q   H   C   P   E   H   R   P   N   F   T   1170

3511  ACT ATT TTA GAG AGG ATC AAC TAC TGT ACA CAG GAT CCT GAT GTC   3555
1171   T   I   L   E   R   I   N   Y   C   T   Q   D   P   D   V   1185

3556  ATC AAC ACC CCT CTT CCA GTG GAG TGT GGT CCT CCT GTT GAA GAA   3600
1186   I   N   T   P   L   P   V   E   C   G   P   P   V   E   E   1200

3601  GAA GGG GGC ACT GTG ATC CGC CCT GAC GGA TCA GGC AGC ATG ACC   3645
1201   E   G   G   T   V   I   R   P   D   G   S   G   S   M   T   1215

3646  CCT CTT CTA GTT GCC CGC TCT TTG TCC CAG GAT GCT TCC CCT CGA   3690
1216   P   L·  L   V   A   R   S   L   S   Q   D   A   S   P   R   1230

3691  GCA AGC ATC ACC AGT GTC ACT CCA CAG GCC CTC AAA CCC CGT TTG   3735
1231   A   S   I   T   S   V   T   P   Q   A   L   K   P   R   L   1245

3736  CAG CTA CAA AGA CCA GTC CAC CTC ACA CAA GAA GTG GGC ACA TAC   3780
1246   Q   L   Q   R   P   V   H   L   T   Q   E   V   G   T   Y   1260

3781  CGA GAG ACA CTG GAG CCC TGT TGG GCA GAG CCT GTT CCT GCT TCA   3825
1261   R   E   T   L   E   P   C   W   A   E   P   V   P   A   S   1275

3826  GGA GTC TGT CCA GGA CCA TGG CTT CAA GTC CCA GAG CAT CGG CCA   3870
1276   G   V   C   P   G   P   W   L   Q   V   P   E   H   R   P   1290

3871  TGC TCC AGA AGT AGT TCA TCA TCG GGC AGC CAG AAG CTA AAG AAC   3915
1291   C   S   R   S   S   S   S   S   G   S   Q   K   L   K   N   1305

3916  AAG ACT AAA AAC CTC TGG AAC CCC ACC TAC GGT TCC TGG GTC CTA   3960
1306   K   T   K   N   L   W   N   P   T   Y   G   S   W   V   L   1320

3961  GAG AGC TTT GGA CGT GGA AAG TCA GCT CTG TGC CAC ACT CAG TCT   4005
1321   E   S   F   G   R   G   K   S   A   L   C   H   T   Q   S   1335
```

```
4006  ATG CCT CTC TCC TGC AAC CCC ACA TCT GTT TCT GCC CCG TCC TCT   4050
1336   M   P   L   S   C   N   P   T   S   V   S   A   P   S   S   1350

4051  ACC TCG GAG CAC ACA GAC CCT GTA GTA GAA GTT AAT GCA AAT GTT   4095
1351   T   S   E   H   T   D   P   V   V   E   V   N   A   N   V   1365

4096  TCA GCT TCC CCA CCT CCA TCT GCT GCA CCC TCA CAA ACT ACG TTG   4140
1366   S   A   S   P   P   P   S   A   A   P   S   Q   T   T   L   1380

4141  ACT CCC ACA GCG GCT CCC TCT AGG AAG AGT CCC ACA GGG GCT GCT   4185
1381   T   P   T   A   A   P   S   R   K   S   P   T   G   A   A   1395

4186  GGT GTC TCA CTT GCC ACT GTT ATG GAC TTA GCA AAG CTT CAG AGC   4230
1396   G   V   S   L   A   T   V   M   D   L   A   K   L   Q   S   1410

4231  TTC CCA TGT GGC AAC GTA AAC TAT GCG TAT GAT GAA CAG AGC TAC   4275
1411   F   P   C   G   N   V   N   Y   A   Y   D   E   Q   S   Y   1425

4276  GAG ACT GAG AGT TTG CCT GTG GTT GTG TCC AAA TCT CTG GAG CCC   4320
1426   E   T   E   S   L   P   V   V   V   S   K   S   L   E   P   1440

4321  AGC ACC AGC TCT GCA GCT ACA TCT TCA CTG GTT GCC CTC AGC CAA   4365
1441   S   T   S   S   A   A   T   S   S   L   V   A   L   S   Q   1455

4366  GCC AGC AGC TTT ACC CAC AAA CCA TTG GTG AAG CGT CAT GCC AGT   4410
1456   A   S   S   F   T   H   K   P   L   V   K   R   H   A   S   1470

4411  TAT GGA CAT GAG GAT GTA AGG AGG TAC ACC CAG CCT GAG AAG CCC   4455
1471   Y   G   H   E   D   V   R   R   Y   T   Q   P   E   K   P   1485

4456  ACC AGG GAC AGG GAT TCT GGC TTT TCC TTG TCT GAA GAC CTG AGT   4500
1486   T   R   D   R   D   S   G   F   S   L   S   E   D   L   S   1500

4501  GTC ACC CCT GTG TAG
1501   V   T   P   V   *
```

4. CLUSTAL W (1.82) multiple sequence alignment showing aminoacid sequences for zebrafish clone1 and clone3 (Z1 and Z3), human ALK (HALK) and mouse ALK (MALK)

```
Z1              MDYITRQTFVKLALFIFTVVRSSCALLEKAAEDPVHPNPLQS-SPAEDSDVS----FCDF 55
Z3              MDYITRQTFVKLALFIFTVVRSSCALLEKAAEDPVHPNPLQS-SPAEDSDVS----FCDF 55
HALK            MGAIG--LLWLLPLLLSTAAVGSGMGTGQRAGSPAAGSPLQPREPLSYSRLQRKSLAVDF 58
MALK            MGAAG--FLWLLPPLLLAAASYSGAATDQRAGSPASGPPLQPREPLSYSRLQRKSLAVDF 58
                *.      :  *. :: :..  *     : * .*.   ***. .* . * :.      **

Z1              ESP------------------------------------------------------- 58
Z3              ESP------------------------------------------------------- 58
HALK            VVPSLFRVYARDLLLPP--SSSELKAGRPEARGSLALDCAPLLRLLGPAPGVSWTAG--S 114
MALK            VVPSLFRVYARDLLLPQPRSPSEPEAGGLEARGSLALDCEPLLRLLGPLPGISWADGASS 118
                  *

Z1              ----CSWTLSSHSTGG------------------------------------------- 70
Z3              ----CSWTLSSHSTGG------------------------------------------- 70
HALK            PAPAEARTLSRVLKGGSVRKLRRAKQLVLELGEEAILEGCVGPPGE-AAVGLLQFNLSEL 173
MALK            PSPEAGPTLSRVLKGGSVRNVRRAKQLVLELGEETILEGCIGPPEEVAAVGILQFNLSEL 178
                     . ***   .**

Z1              -DWFITSAQQHR--------------------SNRRDTQP---IRDYSTGKSEGHFLLLK 106
Z3              -DWFITSAQQHR--------------------SNRRDTQP---IRDYSTGKSEGHFLLLK 106
HALK            FSWWIRQGEGRLRIRLMPEKKASEVGREGRLSAAIRASQPRLLFQIFGTGHSSLESPTNM 233
MALK            FSWWILHGEGRLRIRLMPEKKASEVGREGRLSSAIRASQPRLLFQIFGTGHSSLESPSET 238
                 .*:*  .: :                    :  * :**   :: :.**:*. .

Z1              PSS--------------------SHLSAG--RCSFHMTSPVVLS---------------- 128
Z3              PSS--------------------SHLSAG--RCSFHMTSPVVLS---------------- 128
HALK            PSPSPDYFTWNLTWIMKDSFPFLSHRSRYGLECSFDFPCELEYSPPLHDLRNQSWSWRRI 293
MALK            PSP-PGTFMWNLTWTMKDSFPFLSHRSRYGLECSFDFPCELEYSPPLHNHGNQSWSWRHV 297
                **.                    ** *    .***.:.. :  *

Z1              -SGPFCHLQLARFQPEPHAG----------NISAFVKHT-------DSIDIKPIDLTIKE 170
Z3              -SGPFCHLQLARFQPEPHAG----------NISAFVKHT-------DSIDIKPIDLTIKE 170
HALK            PSEEASQMDLLDGPGAERSKEMPRGSFLLLNTSADSKHTILSPWMRSSSEHCTLAVSVHR 353
MALK            PSEEASRMNLLDGPEAEHSQEMPRGSFLLLNTSADSKHTILSPWMRSSSDHCTLAVSVHR 357
                 *   .:::*       ::           * **  ***       .* :  .: ::::.

Z1              QESDSS-------------------------QWEVLEAVIGQLNEPFQVTVQYSACSSHE 205
Z3              QESDSS-------------------------QWEVLEAVIGQLNEPFQVTVQYSACSSHE 205
HALK            HLQPSGRYIAQLLPHNEAAREILLMPTPGKHGWTVLQGRIGRPDNPFRVALEY--ISSGN 411
MALK            HLQPSGRYVAQLLPHNEAGREILLVPTPGKHGWTVLQGRVGRPANPFRVALEY--ISSGN 415
                : . *.                          * **:. :*:  :**:*:::*   ** :

Z1              VGFLAFDSLELKNCVMGDDYVDLGSDCEKYSSLQCHSGGCIEKQRVCDFHTDCPEGEDEG 265
Z3              VGFLAFDSLELKNCVMGDDYVDLGSDCEKYSSLQCHSGGCIEKQRVCDFHTDCPEGEDEG 265
HALK            RSLSAVDFFALKNCSEG---TSPGSKMALQSSFTCWNGTVLQLGQACDFHQDCAQGEDES 468
MALK            RSLSAVDFFALKNCSEG---TSPGSKMALQSSFTCWNGTVLQLGQACDFHQDCAQGEDEG 472
                 .: *.* : ****  *   .. **.    **: * .*  ::  :.**** **.:****.

Z1              LICSTLPLGSYCSFELGSCGWSAADTQSSWRLVSGQQLIEDTHLLGTTLKNTQGHFLFLK 325
Z3              LICSTLPLGSYCSFELGSCGWSAADTQSSWRLVSGQQLIEDTHLLGTTLKNTQGHFLFLK 325
HALK            QMCRKLPVGFYCNFEDGFCGWTQG------TLSPHTPQWQVRTLKDARFQDHQDHALLLS 522
MALK            QLCSKLPAGFYCNFENGFCGWTQS------PLSPHMPRWQVRTLRDAHSQGHQGRALLLS 526
                 :* .** * **.** * ***: .       * .     :   * .:  :. *.: *:*.

Z1              VRGHGDEREALVQSPALPSTISNQDCQLQFSLYRYGDFNGTVLLSVVES--GASAPALIW 383
Z3              VRGHGDEREALVQSPALPSTISNQDCQLQFSLYRYGDFNGTVLLSVVES--GASAPALIW 383
HALK            TTDVPASESATVTSATFPAPIKSSPCELRMSWLIRGVLRGNVSLVLVENKTGKEQGRMVW 582
MALK            TTDILASEGATVTSATFPAPMKNSPCELRMSWLIRGVLRGNVSLVLVENKTGKEQSRTVW 586
                . .   .. * * *.::*:.:... *:*::*    * :.*.* * :**.  * .    :*
```

```
Z1              ERSG-HWKDAWQEITLPITEILNGFHLKVQAFWTSGSKADIALDDISLSAACFDTELN-- 440
Z3              ERSG-HWKDAWQEITLPITEILNGFHLKVQAFWTSGSKADIALDDISLSAACFDTELN-- 440
HALK            HVAAYEGLSLWQWMVLPLLDVSDRFWLQMVAWWGQGSRAIVAFDNISISLDCYLTISGED 642
MALK            HVATDEGLSLWQHTVLSLLDVTDRFWLQIVTWWGPGSRATVGFDNISISLDCYLTISGEE 646
                .  :   .   . **  .*.: :: : * *:: ::*  **:* :.:*:**:*  *: *  .

Z1              ELLHEGLPHD---LDFSPLPEPSASEASP---------ITWWFTSCGASGPFGPTQAQCD 488
Z3              ELLHEGLPHD---LDFSPLPEPSASEASP---------ITWWFTSCGASGPFGPTQAQCD 488
HALK            KILQNTAPKSRNLFERNPNKELKPGENSPRQTPIFDPTVHWLFTTCGASGPHGPTQAQCN 702
MALK            KMSLNSVPKSRNLFEKNPNKESKSWANISGPTPIFDPTVHWLFTTCGASGPHGPTQAQCN 706
                ::  :  *:.   :: .*  * ..    .         : * **:******.*******:

Z1              SAYRNTNVSVVVGKEGPLRGVQMWKVPATNTYKISAYGAAGGKGAKNHNKRSHGVFISAT 548
Z3              SAYRNTNVSVVVGKEGPLRGVQMWKVPATNTYKISAYGAAGGKGAKNHNKRSHGVFISAT 548
HALK            NAYQNSNLSVEVGSEGPLKGIQIWKVPATDTYSISGYGAAGGKGGKNTMMRSHGVSVLGI 762
MALK            NAYQNSNLSVVVGSEGPLKGVQIWKVPATDTYSISGYGAAGGKGGKNTMMRSHGVSVLGI 766
                .**:*:*:** **.****:*:*:******:**.**.********.**   ***** : .

Z1              FPLEKGDILYILIGHQGEDACPGRNPQTHKICLGESSVIEDGFDSDGSALKWAGGGGGGG 608
Z3              FPLEKGDILYILIGHQGEDACPGRNPQTHKICLGESSVIEDGFDSDGSALKWAGGGGGGG 608
HALK            FNLEKDDMLYILVGQQGEDACPSTNQLIQKVCIGENNVIEEEIRVNRSVHEWAGGGGGGG 822
MALK            FNLEKGDTLYILVGQQGEDACPRANQLIQKVCVGENNVIEEEIRVNRSVHEWAGGGGGGG 826
                * ***.* ****:*:******* *   :*:*:**..***: :  : *. :*********

Z1              GATYIYRMENGQPLPLLIAAGGGGKAYLEDPESSQDQSFREQYENDTTVSGVSGRSGAAG 668
Z3              GATYIYRMENGQPLPLLIAAGGGGKAYLEDPESSQDQSFREQYENDTTVSGVSGRSGAAG 668
HALK            GATYVFKMKDGVPVPLIIAAGGGGRAYG----AKTDTFHPERLENNSSVLGLNGNSGAAG 878
MALK            GATYVFKMKDGVPVPLIIAAGGGGRAYG----AKRETFHPERLESNSSVLGLNGNSGAAG 882
                ****:::*::* *:**:*******:**     :. :  . *: *.:::* *:.*.*****

Z1              GGGGWSDVSSLSWAGKSLVEGGQGGSSCPEALSVLGWATFGGFGGGGGACSAGGGGGGYR 728
Z3              GGGGWSDVSSLSWAGKSLVEGGQGGSSCPEALSVLGWATFGGFGGGGGACSAGGGGGGYR 728
HALK            GGGGWNDNTSLLWAGKSLQEGATGGHSCPQAMKKWGWETRGGFGGGGGGCSSGGGGGGYI 938
MALK            GGGGWNDNTSLLWAGKSLLEGAAGGHSCPQAMKKWGWETRGGFGGGGGGCSSGGGGGGYI 942
                *****.* :** ****** **. ** ***:*:.  ** * ********.**:*******

Z1              GGDAPLLDDISADGQDGLSFVHPMGKIFLQPLAAMESHGEAEIVVYLNCSHCKTQSCKRD 788
Z3              GGDAPLLDDISADGQDGLSFVHPMGKIFLQPLAAMESHGEAEIVVYLNCSHCKTQSCKRD 788
HALK            GGNAASNNDPEMDGEDGVSFISPLGILYTPALKVMEGHGEVNIKHYLNCSHCEVDECHMD 998
MALK            GGNAASNNDPEMDGEDGVSFISPLGILYTPALKVMEGHGEVNIKHYLNCSHCEVDECHMD 1002
                **:*.  :* . **:**:**: *:* ::  .* .**.***.:*  *******:.:.*: *

Z1              EDTKLILCLCDSDEVLAPDNVTCAGTKHSLCQFINKHLQHNSSPLVCPPLVVPMGSLADG 848
Z3              EDTKLILCLCDSDEVLAPDNVTCA-------------------------VVPMGSLADG 822
HALK            PESHKVICFCDHGTVLAEDGVSCI-----------------------------VSPTPEP 1029
MALK            PESHKVICFCDHGTVLADDGVSCI-----------------------------VSPTPEP 1033
                 ::: ::*:** . *** *.*:*                                :.. .:

Z1              PPSLVFIMAVIVSTVVTGVVLTCASLTLIYYRKKNHLHAVRIRLQSPEYKLSKIRSSTIM 908
Z3              PPSLVFIMAVIVSTVVTGVVLTCASLTLIYYRKKNHLHAVRIRLQSPEYKLSKIRSSTIM 882
HALK            HLPLSLILSVVTSALVAALVLAFSGIMIVYRRKHQELQAMQMELQSPEYKLSKLRTSTIM 1089
MALK            HLPLSLILSVVTSALVAALVLAFSGIMIVYRRKHQELQAMQIQLQSPEYKLSKLRTSTIM 1093
                  .* :*::*:.*::*:.:**: :.: ::* **::.*:*:::.**********:*:****

Z1              TDYNPNYGYFGKAASLSELKEVPRKNITLLRALGHGAFGEVYEGQVLGMNGENTAMQVAI 968
Z3              TDYNPNYGYFGKAASLSELKEVPRKNITLLRALGHGAFGEVYEGQVLGMNGENTAMQVAI 942
HALK            TDYNPNYCFAGKTSSISDLKEVPRKNITLIRGLGHGAFGEVYEGQVSGMPNDPSPLQVAV 1149
MALK            TDYNPNYCFAGKTSSISDLKEVPRKNITLIRGLGHGAFGEVYEGQVSGMPNDPSPLQVAV 1153
                ******* : **::*:*:***********:*.************** ** .: :.:***:

Z1              KTLPEICSEQDEMDFLMEALIMSKFSHQNIVRCIGVSLQILPRFILLELMTGGDMKSFLR 1028
Z3              KTLPEICSEQDEMDFLMEALIMSKFSHQNIVRCIGVSLQILPRFILLELMTGGDMKSFLR 1002
HALK            KTLPEVCSEQDELDFLMEALIISKFNHQNIVRCIGVSLQSLPRFILLELMAGGDLKSFLR 1209
MALK            KTLPEVCSEQDELDFLMEALIISKFNHQNIVRCIGVSLQALPRFILLELMAGGDLKSFLR 1213
                *****:******:********:***.************* **********:***:*****
```

```
Z1              LNRPRTNHSSSLSMLELLHMARDIALGCRYLEENHFIHRDIAARNCLLTCPGPDRVAKIG 1088
Z3              LNRPRTNHSSSLSMLELLHMARDIALGCRYLEENHFIHRDIAARNCLLTCPGPDRVAKIG 1062
HALK            ETRPRPSQPSSLAMLDLLHVARDIACGCQYLEENHFIHRDIAARNCLLTCPGPGRVAKIG 1269
MALK            ETRPRPNQPTSLAMLDLLHVARDIACGCQYLEENHFIHRDIAARNCLLTCPGAGRIAKIG 1273
                 .***..:.:**:**:***:***** **:**********************..*:****

Z1              DFGMARDIYRASYYRKGGRAMLPVKWMPPEAFLEGIFTCKTDTWSFGVLLWEIFSLGYMP 1148
Z3              DFGMARDIYRASYYRKGGRAMLPVKWMPPEAFLEGIFTCKTDTWSFGVLLWEIFSLGYMP 1122
HALK            DFGMARDIYRASYYRKGGCAMLPVKWMPPEAFMEGIFTSKTDTWSFGVLLWEIFSLGYMP 1329
MALK            DFGMARDIYRASYYRKGGCAMLPVKWMPPEAFMEGIFTSKTDTWSFGVLLWEIFSLGYMP 1333
                ****************** ************:*****.********************

Z1              YPCKTNQEVLEFVTGGGRMDPPKSCPGPVYRIMTQCWQHCPEHRPNFTTILERINYCTQD 1208
Z3              YPCKTNQEVLEFVTGGGRMDPPKSCPGPVYRIMTQCWQHCPEHRPNFTTILERINYCTQD 1182
HALK            YPSKSNQEVLEFVTSGGRMDPPKNCPGPVYRIMTQCWQHQPEDRPNFAIILERIEYCTQD 1389
MALK            YPSKSNQEVLEFVTSGGRMDPPKNCPGPVYRIMTQCWQHQPEDRPNFAIILERIEYCTQD 1393
                **.*:*********.********.*************** **.****: *****:*****

Z1              PDVINTPLPVECGPPVEEEGGTVIRPDGSGSMTPLLVARSLSQDASPRASITSVTPQALK 1268
Z3              PDVINTPLPVECGPPVEEEGGTVIRPDGSGSMTPLLVARSLSQDASPRASITSVTPQALK 1242
HALK            PDVINTALPIEYGPLVEEEEKVPVRPKDPEGVPPLLVSQQAKREE-------ERSPAAPP 1442
MALK            PDVINTALPIEYGPVVEEEEKVPMRPKDPEGMPPLLVSPQPAKHE-------EAS-AAPQ 1445
                ******.**:* ** ****  . :**... .:.****: .  :.        . :  *

Z1              PRLQLQRPVHLTQEVGTYRETLEPCWAEPVPASGVCPG-PWLQVPEHRPCSRSSSSSGSQ 1327
Z3              PRLQLQRPVHLTQEVGTYRETLEPCWAEPVPASGVCPG-PWLQVPEHRPCSRSSSSSGSQ 1301
HALK            PLPTTSSGKAAKKPTAAEVSVRVPRGP------AVEGGHVNMAFSQSNPPSELHKVHGS- 1495
MALK            PAALTAPGPSVKKPPGAGAGAGAGAGAGPVPRGAADRGHVNMAFSQPNPPPELHKGPGS- 1504
                *          .:  .:   .     .      ..  *   : ..: .* ..  .  **

Z1              KLKNKTKNLWNPTYGSWVLESFGRGKSALCHTQSMPLSCNPTSVSAPSSTSEHTDPVVEV 1387
Z3              KLKNKTKNLWNPTYGSWVLESFGRGKSALCHTQSMPLSCNPTSVSAPSSTSEHTDPVVEV 1361
HALK            --RNKPTSLWNPTYGSWFTE-------------------KPTKKNNPIAKKEPHDRGNLG 1534
MALK            --RNKPTSLWNPTYGSWFTA-------------------KPAKKTHPPPGAEPQARAG-A 1542
                  :**...*********.                     :*:. . * .  *

Z1              NANVSASPPPSAAPSQTTLTPTAAPSRKSPTGAAGVSLATVMDLAKLQSFPCGNVNYAYD 1447
Z3              NANVSASPPPSAAPSQTTLTPTAAPSRKSPTGAAGVSLATVMDLAKLQSFPCGNVNYAYD 1421
HALK            LEGSCTVPPNVATGRLPGASLLLEPSSLTAN-------MKEVPLFRLRHFPCGNVNYGYQ 1587
MALK            AEGGWTGPG--AGPRRAEAALLLEPSALSAT-------MKEVPLFRLRHFPCPNVNYGYQ 1593
                  .  : *   *    .  :    **  :..        . : * :*: *** ****.*:

Z1              EQSYETESLPVVVSKSLEPSTSSAATSSLVALSQASSFTHKPLVKRHASYGHEDVRRYTQ 1507
Z3              EQSYETESLPVVVSKSLEPSTSSAATSSLVALSQASSFTHKPLVKRHASYGHEDVRRYTQ 1481
HALK            QQ-----------GLPLEAATAPGAGHYEDTILKSKNSMNQPGP---------------- 1620
MALK            QQ-----------GLPLEATAAPG-----DTMLKSKNKVTQPGP---------------- 1621
                :*           . .**.::::..      :: ::..   :*

Z1              PEKPTRDRDSGFSLSEDLSVTPV 1530
Z3              PEKPTRDRDSGFSLSEDLSVTPV 1504
HALK            ------------------------
MALK            ------------------------
```

5.CLUSTAL W (1.82) multiple sequence alignment between the aminoacid sequences from human ALK, mouse ALK, human LTK precursor (LHp), human LTK (LH), mouse LTK (LM), mouse LTK precursor (LMp), zebrafish clone 1 and 3 (Z1 and Z3), and drosophila ALK (DALK).

```
HALK            MGAIGLLWLLPLLLSTAAVGSGMGTGQRAGSPAAGSPLQPREPLSYSRLQRKSLAVDFVV 60
MALK            MGAAGFLWLLPPLLLAAASYSGAATDQRAGSPASGPPLQPREPLSYSRLQRKSLAVDFVV 60
LHp             ------------------------------------------------------------
LH              ------------------------------------------------------------
LM              ------------------------------------------------------------
LMp             ------------------------------------------------------------
Z1              ---MDYITRQTFVKLALFIFTVVRSSCALLEKAAEDPVHPNPLQSSPAEDSDVSFCDFES 57
Z3              ---MDYITRQTFVKLALFIFTVVRSSCALLEKAAEDPVHPNPLQSSPAEDSDVSFCDFES 57
DALK            MPQLRQLVVVLAILSLLVDPSFAQRPLIMNHTFSSLPMSQHPSFNSWPSASGGKQLRNGR 60

HALK            PSLFRVYARDLLLPP--SSSELKAGRPEARGSLALDCAPLLRLLGPAPGVSWTAG--SPA 116
MALK            PSLFRVYARDLLLPQPRSPSEPEAGGLEARGSLALDCEPLLRLLGPLPGISWADGASSPS 120
LHp             ------------------------------------------------------------
LH              ------------------------------------------------------------
LM              ------------------------------------------------------------
LMp             ------------------------------------------------------------
Z1              PCSWTLSSHS-------TGGDWFITSAQQHRSNRRDTQPIR---------DYSTGKSEGH 101
Z3              PCSWTLSSHS-------TGGDWFITSAQQHRSNRRDTQPIR---------DYSTGKSEGH 101
DALK            LHEQPLPLPQPTLPVATQYEDYEAAPPGSRRDNKRRLAQMAQRPGSGNGGRRGMESLRKE 120

HALK            PAEARTLSRVLKGGSVRKLRRAKQLVLELGEEAILEGCVGPPGE-AAVGLLQFNLSELFS 175
MALK            PEAGPTLSRVLKGGSVRNVRRAKQLVLELGEETILEGCIGPPEEVAAVGILQFNLSELFS 180
LHp             ------------------------------------------------------------
LH              ------------------------------------------------------------
LM              ------------------------------------------------------------
LMp             ------------------------------------------------------------
Z1              FLLLKPSSSHLSAGRCS-FHMTSPVVLSSGPFCHLQLARFQPEPHAGN----------IS 150
Z3              FLLLKPSSSHLSAGRCS-FHMTSPVVLSSGPFCHLQLARFQPEPHAGN----------IS 150
DALK            LMNPSVGGPGGGISGGGSGYPSYSATSSIGRIDGLGQGLGSADRYGALYDQLRQPSGMAA 180

HALK            WWIRQGEGRLRIRLMPEKKASEVGREGRLSAAIRASQPRLLFQIFGTGHSSLESPTNMPS 235
MALK            WWILHGEGRLRIRLMPEKKASEVGREGRLSSAIRASQPRLLFQIFGTGHSSLESPSETPS 240
LHp             ------------------------------------------------------------
LH              ------------------------------------------------------------
LM              ------------------------------------------------------------
LMp             ------------------------------------------------------------
Z1              AFVKHTDSIDIKPIDLTIKEQE----------SDSSQWEVLEAVIG----QLNEPFQVTV 196
Z3              AFVKHTDSIDIKPIDLTIKEQE----------SDSSQWEVLEAVIG----QLNEPFQVTV 196
DALK            MTAGPLGGPYSTRFPPLYAGVDDQPKRSNRKNSISELYKLKKALNQADEPAGGVTHGNFE 240

HALK            PSPDYFTWNLTWIMKDSFPFLS-------------HRSRYGLECSFDFPCELEYSPPLHD 282
MALK            P-PGTFMWNLTWTMKDSFPFLS-------------HRSRYGLECSFDFPCELEYSPPLHN 286
LHp             ------------------------------------------------------------
LH              ------------------------------------------------------------
LM              ------------------------------------------------------------
LMp             ------------------------------------------------------------
Z1              QYSACSSHEVGFLAFDSLELKN-------------CVMG---DDYVDLGSDCEKYSSLQC 240
Z3              QYSACSSHEVGFLAFDSLELKN-------------CVMG---DDYVDLGSDCEKYSSLQC 240
DALK            GDPQPSSADPLQEIKDRIRDTSPNANTYAPHTDATPSSEYEYELGVKCNFETPCSWTWGN 300
```

```
HALK        LRNQSWSWRRIPSEEASQMDLLDGPGAERSKEMPRGSFLLLNTSADSKHTILSPWMRSSS 342
MALK        HGNQSWSWRHVPSEEASRMNLLDGPEAEHSQEMPRGSFLLLNTSADSKHTILSPWMRSSS 346
LHp         ------------------------------------------------------------
LH          ------------------------------------------------------------
LM          ------------------------------------------------------------
LMp         ------------------------------------------------------------
Z1          HSGGCIEKQRVCDFHTDCPEGEDEGLICSTLPLGSYCSFELGSCGWSAADTQSSWRLVSG 300
Z3          HSGGCIEKQRVCDFHTDCPEGEDEGLICSTLPLGSYCSFELGSCGWSAADTQSSWRLVSG 300
DALK        YSDGFQVITGTELSKRNLTGLLPGPAADSIDDANGHFLYARVNPSSRPLNLTSPEFSTTM 360

HALK        EHCTLAVSVH-RHLQP-SGRYIAQLLPHNEAAREILLMPTPGKHGWTVLQGRIGRPDNPF 400
MALK        DHCTLAVSVH-RHLQP-SGRYVAQLLPHNEAGREILLVPTPGKHGWTVLQGRVGRPANPF 404
LHp         ------------------------------------------------------------
LH          ------------------------------------------------------------
LM          ------------------------------------------------------------
LMp         ------------------------------------------------------------
Z1          QQLIEDTHLLGTTLKN-TQGHFLFLKVRGHGDEREALVQSPALPSTISNQD--------- 350
Z3          QQLIEDTHLLGTTLKN-TQGHFLFLKVRGHGDEREALVQSPALPSTISNQD--------- 350
DALK        EKCFLEVYMHQSDMSHGLSRVVVELLHTAESSWVPAEILGDNVRQWTRKVYRLGRVSRDF 420

HALK        RVALEYISSGNRSLSAVDFFALKNCS-----EGTSPGSKMALQSSFTCWNGTVLQLGQAC 455
MALK        RVALEYISSGNRSLSAVDFFALKNCS-----EGTSPGSKMALQSSFTCWNGTVLQLGQAC 459
LHp         ---------------------------------------------MGCWGQLLVWFGAAG 15
LH          ---------------------------------------------MGCWGQLLVWFGAAG 15
LM          ------------------------------------------------------------
LMp         ---------------------------------------------MGCSHRLLLWLGAAG 15
Z1          -CQLQFSLYRYGDFNGTVLLSVVESG-----ASAPALIWERSGHWKDAWQEITLPITEIL 404
Z3          -CQLQFSLYRYGDFNGTVLLSVVESG-----ASAPALIWERSGHWKDAWQEITLPITEIL 404
DALK        RIVFEVVPDLRVGQKGHVALDNLRMVNCFPEGTKSEKCSTSQVKCTSSKVPVCIHLPRIC 480

HALK        DFHQDCAQGEDESQMCRKLPVGFYCNFEDGFCGWTQGTLSPHTPQWQVRTLKDARFQDHQ 515
MALK        DFHQDCAQGEDEGQLCSKLPAGFYCNFENGFCGWTQSPLSPHMPRWQVRTLRDAHSQGHQ 519
LHp         ------------AILCSSPGS------------------------------------- 24
LH          ------------AILCSSPGS------------------------------------- 24
LM          ------------------------------------------------------------
LMp         ------------TILCSNSEF------------------------------------- 24
Z1          N------------GFHLKVQAFWTS---------------------------------- 417
Z3          N------------GFHLKVQAFWTS---------------------------------- 417
DALK        DITRDCDEAEDEQQSCDKIPYGGRCDFEEDWCGWRDSGKTTLTWSRHTGSSPTHDTGPDG 540

HALK        DHALLLSTTDVPASESATVTSATFPAPIKSSPCELRMSWLIRGVLRGNVSLVLVENKTGK 575
MALK        GRALLLSTTDILASEGATVTSATFPAPMKNSPCELRMSWLIRGVLRGNVSLVLVENKTGK 579
LHp         -QETFLRSSPLPLASPSPQDPKVSAPPSILEPASP-------------------LNSPGT 64
LH          -QETFLRSSPLPLASPSPRDPKVSAPPSILEPASP-------------------LNSPGT 64
LM          ------------------------------------------------------------
LMp         -QAPFLTPSLLPVLVLNSQEQKVTPTPSKLEPASL-------------------PNPLGT 64
Z1          GSKADIALDDISLS-AACFDTELNELLHEGLPHDLDFS---------------PLPEPSAS 462
Z3          GSKADIALDDISLS-AACFDTELNELLHEGLPHDLDFS---------------PLPEPSAS 462
DALK        DHTMQHLQNNTSGYYMLVNMNQHMNNSEKNSIIGFASNAIMVSKTFNPPPSVHGNPDSPY 600

HALK        EQGRMVWHVAAYEGLSLWQWMVLPLLDVSDRFWLQMVAWWGQGSRAIVAFDNISISLDCY 635
MALK        EQSRTVWHVATDEGLSLWQHTVLSLLDVTDRFWLQIVTWWGPGSRATVGFDNISISLDCY 639
LHp         EGS------------------------------------------------------- 67
LH          EGS------------------------------------------------------- 67
LM          ------------------------------------------------------------
LMp         RGP------------------------------------------------------- 67
Z1          EAS------------------------------------------------------- 465
Z3          EAS------------------------------------------------------- 465
DALK        RNSCVVRFFIHQFGKNPGSINLSVVEMKEKENITTTLWWSTKNQGSDWMRAEYVLPNITS 660
```

```
HALK          LTISGEDKILQNTAPKSRNLFERNPNKELK-----------------PGENSPRQTPIFD 678
MALK          LTISGEEKMSLNSVPKSRNLFEKNPNKESK-----------------SWANISGPTPIFD 682
LHp           ------------------------------------------------------------
LH            ------------------------------------------------------------
LM            ------------------------------------------------------------
LMp           ------------------------------------------------------------
Z1            ------------------------------------------------------------
Z3            ------------------------------------------------------------
DALK          KYYLQFEARMGMRIYSDVAVDDFSLSPECFGLNIPEDHLGGYNYWDVRQNLKSPTYKDFE 720

HALK          PTVHWLFTTCGASGPHGPTQAQCNNAYQNSN------LSVEVGSEGPLKGIQIWKVPATD 732
MALK          PTVHWLFTTCGASGPHGPTQAQCNNAYQNSN------LSVVVGSEGPLKGVQIWKVPATD 736
LHp           ----WLFSTCGASGRHGPTQTQCDGAYAGTS------VVVTVGAAGQLRGVQLWRVPGPG 117
LH            ----WLFSTCGASGRHGPTQTQCDGAYAGTS------VVVTVGAAGQLRGVQLWRVPGPG 117
LM            ------------------------------------------------------------
LMp           ----WVFNTCGASGRSGPTQTQCDGAYTGSS------VMVTVGAAGPLKGVQLWRAPDTG 117
Z1            -PITWWFTSCGASGPFGPTQAQCDSAYRNTN------VSVVVGKEGPLRGVQMWKVPATN 518
Z3            -PITWWFTSCGASGPFGPTQAQCDSAYRNTN------VSVVVGKEGPLRGVQMWKVPATN 518
DALK          YTNYLELTTCDTRGMIGPSQAQCEAAYREQNKTHVLREVHVVEDQSSYKGMQKWKVPHEG 780

HALK          TYSISGYGAAGGKGGKNTMMRSHGVSVLGIFNLEKDDMLYILVGQQGEDACPSTNQLIQK 792
MALK          TYSISGYGAAGGKGGKNTMMRSHGVSVLGIFNLEKGDTLYILVGQQGEDACPRANQLIQK 796
LHp           QYLISAYGAAGGKGAKNHLSRAHGVFVSAIFSLGLGESLYILVGQQGEDACPGGSPESQL 177
LH            QYLISAYGAAGGKGAKNHLSRAHGVFVSAIFSLGLGESLYILVGQQGEDACPGGSPESQL 177
LM            ------------------------------------------------------------
LMp           QYLISAYGAAGGKGAQNHLSRAHGIFLSAVFFLRRGEPVYILVGQQGQDACPGGSPESQL 177
Z1            TYKISAYGAAGGKGAKNHNKRSHGVFISATFPLEKGDILYILIGHQGEDACPGRNPQTHK 578
Z3            TYKISAYGAAGGKGAKNHNKRSHGVFISATFPLEKGDILYILIGHQGEDACPGRNPQTHK 578
DALK          HYTIIAKGASGGLGS-GGVGSSRGSVAVAILELHKNEELYFLVGQQGENACIKSMGVLKE 839

HALK          VCIGENNVIEEEIRVNRS--------VHEWAGGGGGGGGATYVFKMKDGVPVPLIIAAGG 844
MALK          VCVGENNVIEEEIRVNRS--------VHEWAGGGGGGGGATYVFKMKDGVPVPLIIAAGG 848
LHp           VCLGESRAVEEHAAMDGSEG---VPGSRRWAGGGGGGGGATYVFRVRAGELEPLLVAAGG 234
LH            VCLGESRAVEEHAAMDGSEG---VPGSRRWAGGGGGGGGATYVFRLRAGELEPLLVAAGG 234
LM            ------------------------------------------------------------
LMp           VCLGES---GEHATTYGTER---IPGWRRWAGGGGGGGGATSIFRLRAGEPEPLLVAAGG 231
Z1            ICLGESSVIEDGFDSDGS--------ALKWAGGGGGGGGATYIYRMENGQPLPLLIAAGG 630
Z3            ICLGESSVIEDGFDSDGS--------ALKWAGGGGGGGGATYIYRMENGQPLPLLIAAGG 630
DALK          AGCGTDHDLDLAQYSFRSKQDMVKNIYIENGAGGGGGGSYVFLLNQAKNEAVPLLVAGGG 899

HALK          GGRAYGAKTDT----FHPERLENNSSVLGLNGN------SGAAGGGGGWNDNTSLLWAGK 894
MALK          GGRAYGAKRET----FHPERLESNSSVLGLNGN------SGAAGGGGGWNDNTSLLWAGK 898
LHp           GGRAYLRPRDRGRTQASPEKLENRSEAPGSGGR------GGAAGGGGGWTSRAPSPQAGR 288
LH            GGRAYLRPRDRGRTQASPEKLENRSEAPGSGGR------GGAAGG--------------- 273
LM            --------------------METRAAAPGSGGR------GGAAGG--------------- 19
LMp           GGRSYRRRPDRGRTQAVPERLETRAAAPGSGGR------GGAAGGGSGWTSRAHSPQAGR 285
Z1            GGKAYLEDPESSQDQSFREQYENDTTVSGVSGR------SGAAGGGGGWSDVSSLSWAGK 684
Z3            GGKAYLEDPESSQDQSFREQYENDTTVSGVSGR------SGAAGGGGGWSDVSSLSWAGK 684
DALK          GGLGIGQYIDEDFQHGQKAKPLQAPESGQINGEPLNKKTAGPGGGWRAKEDQALSPTYGA 959
                                    .     .*.      .*..**

HALK          SLQEGATGGHSCPQAMKKWGWETR----GGFGGGGGGCSSGGGGGGYIGGNAASNNDPEM 950
MALK          SLLEGAAGGHSCPQAMKKWGWETR----GGFGGGGGGCSSGGGGGGYIGGNAASNNDPEM 954
LHp           SLQEGAEGGQGCSEAWATLGWAAA----GGFGGGGGACTAGGGGGGYRGGDASETDNLWA 344
LH            ------------------------------------------------DASETDNLWA 283
LM            ------------------------------------------------DTSESDLLWA 29
LMp           SPREGAEGGEGCAEAWAALRWAAA----GGFGGGGGACAAGGGGGGYRGGDTSESDLLWA 341
Z1            SLVEGGQGGSSCPEALSVLGWATF----GGFGGGGGACSAGGGGGGYRGGDAPLLDDISA 740
Z3            SLVEGGQGGSSCPEALSVLGWATF----GGFGGGGGACSAGGGGGGYRGGDAPLLDDISA 740
DALK          ALLQGGRGGHSCYVELADNGTSVHRHGQGGFGGGGGGCNTGGGGGGYAGGDVYLTESNGE 1019
                                                                :.   :
```

```
HALK            DGEDGVSFISPLGILYTPALKVMEGHGEVNIKHYLNCSHCEVDECHMDPESHKVICFCDH 1010
MALK            DGEDGVSFISPLGILYTPALKVMEGHGEVNIKHYLNCSHCEVDECHMDPESHKVICFCDH 1014
LHp             DGEDGVSFIHPSSELFLQPLAVTENHGEVEIRRHLNCSHCPLRDCQWQAELQLAECLCPE 404
LH              DGEDGVSFIHPSSELFLQPLAVTENHGEVEIRRHLNCSHCPLRDCQWQAELQLAECLCPE 343
LM              DGEDGTSFVHPSGELYLQPLAVTEGHGEVEIRKHPNCSHCPFKDCQWQAELWTAECTCPE 89
LMp             DGEDGTSFVHPSGELYLQPLAVTEGHGEVEIRKHPNCSHCPFKDCQWQAELWTAECTCPE 401
Z1              DGQDGLSFVHPMGKIFLQPLAAMESHGEAEIVVYLNCSHCKTQSCKRDEDTKLILCLCDS 800
Z3              DGQDGLSFVHPMGKIFLQPLAAMESHGEAEIVVYLNCSHCKTQSCKRDEDTKLILCLCDS 800
DALK            GGSSYISPSRSLREISEIHAGASSGPGAIIIIPAIEGCGCDYRCVALDEFRSKVRCICPD 1079
                .*..  *   .   :      . .. *   *    : . *       :       * *


HALK            GTVLAEDGVSCI--------------------------VSPT---PEPHLPLSLILSVVT 1041
MALK            GTVLADDGVSCI--------------------------VSPT---PEPHLPLSLILSVVT 1045
LHp             GMELAVDNVTCM--------------------------DLHK---PPGPL---VLMVAVV 432
LH              GMELAVDNVTCM--------------------------DLHK---PPGPL---VLMVAVV 371
LM              GTELAVDNVTCM--------------------------DLPT---TASPL---ILMGAVV 117
LMp             GTELAVDNVTCM--------------------------DLPT---TASPL---ILMGAVV 429
Z1              DEVLAPDNVTCAGTKHSLCQFINKHLQHNSSPLVCPPLVVPMGSLADGPPSLVFIMAVIV 860
Z3              DEVLAPDNVTCA--------------------------VVPMGSLADGPPSLVFIMAVIV 834
DALK            GWSLKRDNHTACEIR-----------------------------EEAGKSSFQYLVSILM 1110
                .  *  *. :.                                          ::  :

HALK            SALVAALVLAFSGIMIVYRRKHQELQAMQME--LQSPEYKLSKLR---TSTIMTDYNPNY 1096
MALK            SALVAALVLAFSGIMIVYRRKHQELQAMQIQ--LQSPEYKLSKLR---TSTIMTDYNPNY 1100
LHp             ATSTLSLLMVCGVLILVKQKKWQGLQEMR----LPSPELELSKLR---TSAIRTAPNPYY 485
LH              ATSTLSLLMVCGVLILVKQKKWQGLQEMR----LPSPELELSKLR---TSAIRTAPNPYY 424
LM              AALALSLLMMCAVLILVNQK-CQGLWGTR----LPGPELELSKLR---SSAIRTAPNPYY 169
LMp             AALALSLLMMCAVLILVNQK-CQGLWGTR----LPGPELELSKLR---SSAIRTAPNPYY 481
Z1              STVVTGVVLTCASLTLIYYRKKNHLHAVRIR--LQSPEYKLSKIR---SSTIMTDYNPNY 915
Z3              STVVTGVVLTCASLTLIYYRKKNHLHAVRIR--LQSPEYKLSKIR---SSTIMTDYNPNY 889
DALK            ISLAVLFICIAALIFMLYNRYQRKKQSKKRHKMLVEQDLQLTRLRNNIDDSNLNNFNPNY 1170
                 : .  .:   . : ::  :  .    :    *   : :*:::*    .:  .  ** *

HALK            CFAGKTSSIS-----DLKEVPRKNITLIRGLGHGAFGEVYEGQVSGMPNDPSPLQVAVKT 1151
MALK            CFAGKTSSIS-----DLKEVPRKNITLIRGLGHGAFGEVYEGQVSGMPNDPSPLQVAVKT 1155
LHp             CQVGLGPAQSWPLPPGVTEVSPANVTLLRALGHGAFGEVYEGLVIGLPGDSSPLQVAIKT 545
LH              CQVGLGPAQSWPLPPGVTEVSPANVTLLRALGHGAFGEVYEGLVIGLPGDSSPLQVAIKT 484
LM              CQVGLSPAQPWPLPPGLTEVSPANVTLLRALGHGAFGEVYEGLVTGLPGDSSPLPVAIKT 229
LMp             CQVGLSPAQPWPLPPGLTEVSPANVTLLRALGHGAFGEVYEGLVTGLPGDSSPLPVAIKT 541
Z1              GYFGKAASLS-----ELKEVPRKNITLLRALGHGAFGEVYEGQVLGMNGENTAMQVAIKT 970
Z3              GYFGKAASLS-----ELKEVPRKNITLLRALGHGAFGEVYEGQVLGMNGENTAMQVAIKT 944
DALK            GCDGILNGHIDVN--SLPQVARDSLQLVNALGKGAFGEVYMALYRHRDGDAVEMGVAVKT 1228
                   *   .        : :*.  .: *:..**:******* .      .:   : **:**

HALK            LPEVCSEQDELDFLMEALIISKFNHQNIVRCIGVSLQSLPRFILLELMAGGDLKSFLRET 1211
MALK            LPEVCSEQDELDFLMEALIISKFNHQNIVRCIGVSLQALPRFILLELMAGGDLKSFLRET 1215
LHp             LPELCSPQDELDFLMEALIISKFRHQNIVRCVGLSLRATPRLILLELMSGGDMKSFLRHS 605
LH              LPELCSPQDELDFLMEALIISKFRHQNIVRCVGLSLRATPRLILLELMSGGDMKSFLRHS 544
LM              LPELCSHQDELDFLMEALIISKFSHQNIVRCVGLSFRSAPRLILLELMSGGDMKSFLRHS 289
LMp             LPELCSHQDELDFLMEALIISKFSHQNIVRCVGLSFRSAPRLILLELMSGGDMKSFLRHS 601
Z1              LPEICSEQDEMDFLMEALIMSKFSHQNIVRCIGVSLQILPRFILLELMTGGDMKSFLRLN 1030
Z3              LPEICSEQDEMDFLMEALIMSKFSHQNIVRCIGVSLQILPRFILLELMTGGDMKSFLRLN 1004
DALK            LREDPKREKEEDFLKEAAIMAKFNHPNMVHLIGVCFDRQPYYIVLELLAGGDLQKFLREN 1288
                * *  . :.* *** ** *::** * *:*: :*:.:   *  *:***::***::.*** .

HALK            RPRPSQPSSLAMLDLLHVARDIACGCQYLEENHFIHRDIAARNCLLTCPGPGRVAKIGDF 1271
MALK            RPRPNQPTSLAMLDLLHVARDIACGCQYLEENHFIHRDIAARNCLLTCPGAGRIAKIGDF 1275
LHp             RPHLGQPSPLVMRDLLQLAQDIAQGCHYLEENHFIHRDIAARNCLLSCAGPSRVAKIGDF 665
LH              RPHLGQPSPLVMRDLLQLAQDIAQGCHYLEENHFIHRDIAARNCLLSCAGPSRVAKIGDF 604
LM              RPHPGQLAPLTMQDLLQLAQDIAQGCHYLEENHFIHRDIAARNCLLSCSGASRVAKIGDF 349
LMp             RPHPGQLAPLTMQDLLQLAQDIAQGCHYLEENHFIHRDIAARNCLLSCSGASRVAKIGDF 661
Z1              RPRTNHSSSLSMLELLHMARDIALGCRYLEENHFIHRDIAARNCLLTCPGPDRVAKIGDF 1090
Z3              RPRTNHSSSLSMLELLHMARDIALGCRYLEENHFIHRDIAARNCLLTCPGPDRVAKIGDF 1064
DALK            RNTPERPSLLTMKDLLFCALDVAKGCRYMESKRFIHRDIAARNCLLSSKGPGRVVKIADF 1348
                *    : : * * :**  * *:* **:*:*.::************:. *..*:.**.**
```

```
HALK            GMARDIYRASYYRKGGCAMLPVKWMPPEAFMEGIFTSKTDTWSFGVLLWEIFSLGYMPYP 1331
MALK            GMARDIYRASYYRKGGCAMLPVKWMPPEAFMEGIFTSKTDTWSFGVLLWEIFSLGYMPYP 1335
LHp             GMARDIYRASYYRRGDRALLPVKWMPPEAFLEGIFTSKTDSWSFGVLLWEIFSLGYMPYP 725
LH              GMARDIYRASYYRRGDRALLPVKWMPPEAFLEGIFTSKTDSWSFGVLLWEIFSLGYMPYP 664
LM              GMARDIYQASYYRKGGRTLLPVKWMPPEALLEGLFTSKTDSWSFGVLLWEIFSLGYMPYP 409
LMp             GMARDIYQASYYRKGGRTLLPVKWMPPEALLEGLFTSKTDSWSFGVLLWEIFSLGYMPYP 721
Z1              GMARDIYRASYYRKGGRAMLPVKWMPPEAFLEGIFTCKTDTWSFGVLLWEIFSLGYMPYP 1150
Z3              GMARDIYRASYYRKGGRAMLPVKWMPPEAFLEGIFTCKTDTWSFGVLLWEIFSLGYMPYP 1124
DALK            GMSRDIYRSDYYRKGGKAMLPIKWMPPEAFLDGIFTSKTDVWSFGILLWEVFSLGRSPYP 1408
                **:****::.***:*. ::**:*******:::*:**.*** ****:****:****  ***


HALK            SKSNQEVLEFVTSGGRMDPPKNCPGPVYRIMTQCWQHQPEDRPNFAIILERIEYCTQDPD 1391
MALK            SKSNQEVLEFVTSGGRMDPPKNCPGPVYRIMTQCWQHQPEDRPNFAIILERIEYCTQDPD 1395
LHp             GRTNQEVLDFVVGGGRMDPPRGCPGPVYRIMTQCWQHEPELRPSFASILERLQYCTQDPD 785
LH              GRTNQEVLDFVVGGGRMDPPRGCPGPVYRIMTQCWQHEPELRPSFASILERLQYCTQDPD 724
LM              GHTNQEVLDFIATGNRMDPPRNCPGPVYRIMTQCWQHQPELRPDFGSILERIQYCTQDPD 469
LMp             GHTNQEVLDFIATGNRMDPPRNCPGPVYRIMTQCWQHQPELRPDFGSILERIQYCTQDPD 781
Z1              CKTNQEVLEFVTGGGRMDPPKSCPGPVYRIMTQCWQHCPEHRPNFTTILERINYCTQDPD 1210
Z3              CKTNQEVLEFVTGGGRMDPPKSCPGPVYRIMTQCWQHCPEHRPNFTTILERINYCTQDPD 1184
DALK            GQHNTQVMELVVRGGRLGSPTECPVSIYKVMADCWNPTPEDRPTFITLLEHLTACTQDAS 1468
                 : * :*:::::. *.*:..*  ** .:*::*::**:  ** ** *  :**::  ****..

HALK            VINTALPIEYGPLVEEEEKVPVRPKDPEGVPPLLVSQQAKREEERSPAAPPPLPTTSSGK 1451
MALK            VINTALPIEYGPVVEEEEKVPMRPKDPEGMPPLLVSPQPAKHEEAS-AAPQPAALTAPGP 1454
LHp             VLNSLLPMELGPTPEEEGTSGLGNRSLECLR-----------------PPQPQELSP--E 826
LH              VLNSLLPMELGPTPEEEGTSGLGNRSLECLR-----------------PPQPQELSP--E 765
LM              VLNSPLPMEPGPILEEEEASRLGNRSLEGLR-----------------SPKPLELSS--Q 510
LMp             VLNSPLPMEPGPILEEEEASRLGNRSLEGLR-----------------SPKPLELSS--Q 822
Z1              VINTPLPVECGPPVEEEGGTVIRPDGSGSMTPLLVAR-----SLSQDASPRASITSVTPQ 1265
Z3              VINTPLPVECGPPVEEEGGTVIRPDGSGSMTPLLVAR-----SLSQDASPRASITSVTPQ 1239
DALK            IMNAPLPNILGPTASERDDTVIRPPNGEEFCLAVPDYLVP----LPPGGSNNPSMASGSG 1524
                ::*: **   **  .*.    :   .   .                    .     :

HALK            AAKKPTAAEVSVRVPRGP------AVEGGHVNMAFSQSNPPSELHKVHGS---------- 1495
MALK            SVKKPPGAGAGAGAGAGAGPVPRGAADRGHVNMAFSQPNPPPELHKGPGS---------- 1504
LHp             KLKSWGGSPLGP--------------------------WLSSGLKPLKSR---------- 850
LH              KLKSWGGSPLGP--------------------------WLSSGLKPLKSR---------- 789
LM              NLKSWGGGLLGS--------------------------WLPSGLKTLKPR---------- 534
LMp             NLKSWGGGLLGS--------------------------WLPSGLKTLKPR---------- 846
Z1              ALKPRLQLQRPVHLTQEVGTYRETLEPCWAEPVPASGVCPGPWLQVPEHRPCSRSSSSSG 1325
Z3              ALKPRLQLQRPVHLTQEVGTYRETLEPCWAEPVPASGVCPGPWLQVPEHRPCSRSSSSSG 1299
DALK            YVPELQRQQMSS--------------------------CTPPAVTSPAAP---------- 1548
                                                       . :

HALK            ----RNKPTSLWNPTYGSWFTE------------------KPTKKNNPIAKKEPHDRGN 1532
MALK            ----RNKPTSLWNPTYGSWFTA------------------KPAKKTHPPPGAEPQARAG 1541
LHp.            ----GLQPQNLWNPTYRS---------------------------------------- 864
LH              ----GLQPQNLWNPTYRS---------------------------------------- 803
LM              ----CLQPQNIWNPTYGSWTPR------------------GPQGEDTGIEQCNGSSSSS 571
LMp             ----CLQPQNIWNPTYGSWTPR------------------GPQGEDTGIEQCNGSSSSS 883
Z1              SQKLKNKTKNLWNPTYGSWVLESFGRGKSALCHTQSMPLSCNPTSVSAPSSTSEHTDPVV 1385
Z3              SQKLKNKTKNLWNPTYGSWVLESFGRGKSALCHTQSMPLSCNPTSVSAPSSTSEHTDPVV 1359
DALK            ----HPRPVENIAPTSGDCWETS----------------FILPNSKVEPPVVGSGHDVPL 1588
                      :. .   **  .

HALK            LGLEGSCTVPPNVATGRLPGASLLLEPSSLTAN-------MKEVPLFRLRHFPCGNVNYG 1585
MALK            -AAEGGWTGPG--AGPRRAEAALLLEPSALSAT-------MKEVPLFRLRHFPCPNVNYG 1591
LHp             ------------------------------------------------------------
LH              ------------------------------------------------------------
LM              IPGIQ---------------------------------------------------- 576
LMp             IPGIQ---------------------------------------------------- 888
Z1              EVNANVSASPPPSAAPSQTTLTPTAAPSRKSPTGAAGVSLATVMDLAKLQSFPCGNVNYA 1445
Z3              EVNANVSASPPPSAAPSQTTLTPTAAPSRKSPTGAAGVSLATVMDLAKLQSFPCGNVNYA 1419
DALK            AGGEEAKLISLDTPQPTPTTIQPPLSFASQLDG-------ITLDPSALTKSLPNGNAKQS 1641
```

```
HALK            YQQQGLP-----------LEAATAPGAGHYEDTILKSKNSMNQPGP-------------- 1620
MALK            YQQQGLP-----------LEATAAPG-----DTMLKSKNKVTQPGP-------------- 1621
LHp             ------------------------------------------------------------
LH              ------------------------------------------------------------
LM              ------------------------------------------------------------
LMp             ------------------------------------------------------------
Z1              YDEQSYETESLPVVVSKSLEPSTSSAATSSLVALSQASSFTHKPLVKRHASYGHEDVRRY 1505
Z3              YDEQSYETESLPVVVSKSLEPSTSSAATSSLVALSQASSFTHKPLVKRHASYGHEDVRRY 1479
DALK            YANVQVKNEAEAKVINGVVGNGTVMNGDVSTGSGGADSPFTIQGYAERYKDNNNHSEISC 1701
```

```
HALK            -------------------------
MALK            -------------------------
LHp             -------------------------
LH              -------------------------
LM              -------------------------
LMp             -------------------------
Z1              TQPEKPTRDRDSGFSLSEDLSVTPV 1530
Z3              TQPEKPTRDRDSGFSLSEDLSVTPV 1504
DALK            -------------------------
```

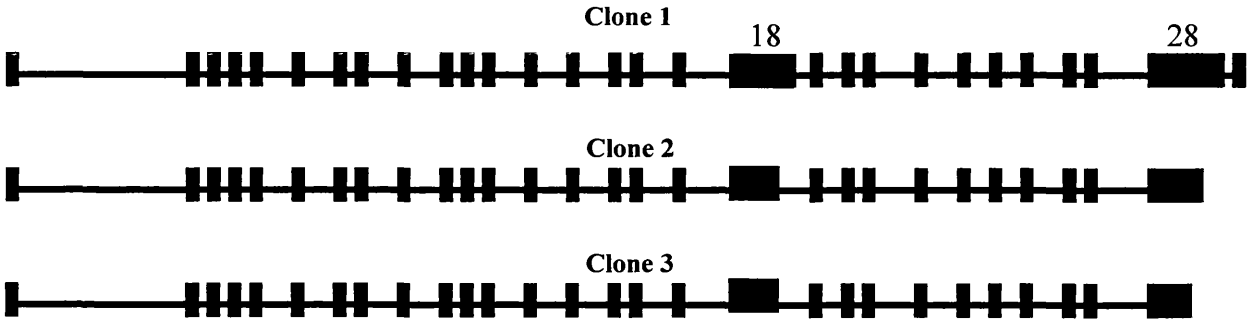


Figure A1. Diagram of the *shd/ALK* genomic structure. Clones 1-3 represent the 3 cDNA clones isolated. Clone 1 has a larger exon 18 because it has more 78 nucleotides than clones 2 and 3. Clone 1 also has a larger exon 28 and an extra exon 29 compared to the other clones. Clones 2 and 3 probably represent the same transcript which was truncated at different positions of exon 28. Detailed information on the sizes and positions of exons/introns is provided in Table 4.1.

References

Agarwal, S., Ramanathan, U. and Naresh, K. N. (2002). Epstein-Barr virus association and ALK gene expression in anaplastic large-cell lymphoma. *Hum Pathol* **33**, 146-52.

Amemiya, C. T., Zhong, T. P., Silverman, G. A., Fishman, M. C. and Zon, L. I. (1999). Zebrafish YAC, BAC, and PAC genomic libraries. *Methods Cell Biol* **60**, 235-58.

Amemiya, C. T. and Zon, L. I. (1999). Generation of a zebrafish P1 artificial chromosome library. *Genomics* **58**, 211-3.

Anderson, D. M., Lyman, S. D., Baird, A., Wignall, J. M., Eisenman, J., Rauch, C., March, C. J., Boswell, H. S., Gimpel, S. D., Cosman, D. et al. (1990). Molecular cloning of mast cell growth factor, a hematopoietin that is active in both membrane bound and soluble forms. *Cell* **63**, 235-43.

Andrews, N. C. (2000). Iron homeostasis: insights from genetics and animal models. *Nat Rev Genet* **1**, 208-17.

Au, W. Y. and Liang, R. (2002). Peripheral T-cell lymphoma. *Curr Oncol Rep* **4**, 434-42.

Aybar, M. J., Glavic, A. and Mayor, R. (2002). Extracellular signals, cell interactions and transcription factors involved in the induction of the neural crest cells. *Biol Res* **35**, 267-75.

Aybar, M. J. and Mayor, R. (2002). Early induction of neural crest cells: lessons learned from frog, fish and chick. *Curr Opin Genet Dev* **12**, 452-8.

Bagnara, J. T., Matsumoto, J., Ferris, W., Frost, S. K., Turner, W. A., Jr., Tchen, T. T. and Taylor, J. D. (1979). Common origin of pigment cells. *Science* **203**, 410-5.

Bai, R. Y., Ouyang, T., Miething, C., Morris, S. W., Peschel, C. and Duyster, J. (2000). Nucleophosmin-anaplastic lymphoma kinase associated with anaplastic large-cell lymphoma activates the phosphatidylinositol 3-kinase/Akt antiapoptotic signaling pathway. *Blood* **96**, 4319-27.

Baker, C. V. and Bronner-Fraser, M. (1997). The origins of the neural crest. Part I: embryonic induction. *Mech Dev* **69**, 3-11.

Barbacid, M. (1993). Nerve growth factor: a tale of two receptors. *Oncogene* **8**, 2033-42.

Baroffio, A., Dupin, E. and Le Douarin, N. M. (1988). Clone-forming ability and differentiation potential of migratory neural crest cells. *Proc Natl Acad Sci U S A* **85**, 5325-9.

Baroffio, A., Dupin, E. and Le Douarin, N. M. (1991). Common precursors for neural and mesectodermal derivatives in the cephalic neural crest. *Development* **112**, 301-5.

Baynash, A. G., Hosoda, K., Giaid, A., Richardson, J. A., Emoto, N., Hammer, R. E. and Yanagisawa, M. (1994). Interaction of endothelin-3 with endothelin-B receptor is

essential for development of epidermal melanocytes and enteric neurons. *Cell* **79**, 1277-85.

Beckmann, G. and Bork, P. (1993). An adhesive domain detected in functionally diverse receptors. *Trends Biochem Sci* **18**, 40-1.

Ben-Neriah, Y. and Bauskin, A. R. (1988). Leukocytes express a novel gene encoding a putative transmembrane protein-kinase devoid of an extracellular domain. *Nature* **333**, 672-6.

Berger, W., Meindl, A., van de Pol, T. J., Cremers, F. P., Ropers, H. H., Doerner, C., Monaco, A., Bergen, A. A., Lebo, R., Warburgh, M. et al. (1992). Isolation of a candidate gene for Norrie disease by positional cloning. *Nat Genet* **2**, 84.

Boldog, F. L., Gemmill, R. M., Wilke, C. M., Glover, T. W., Nilsson, A. S., Chandrasekharappa, S. C., Brown, R. S., Li, F. P. and Drabkin, H. A. (1993). Positional cloning of the hereditary renal carcinoma 3;8 chromosome translocation breakpoint. *Proc Natl Acad Sci U S A* **90**, 8509-13.

Bridge, J. A., Kanamori, M., Ma, Z., Pickering, D., Hill, D. A., Lydiatt, W., Lui, M. Y., Colleoni, G. W., Antonescu, C. R., Ladanyi, M. et al. (2001). Fusion of the ALK gene to the clathrin heavy chain gene, CLTC, in inflammatory myofibroblastic tumor. *Am J Pathol* **159**, 411-5.

Bronner-Fraser, M. and Fraser, S. (1989). Developmental potential of avian trunk neural crest cells in situ. *Neuron* **3**, 755-66.

Brownlie, A., Donovan, A., Pratt, S. J., Paw, B. H., Oates, A. C., Brugnara, C., Witkowska, H. E., Sassa, S. and Zon, L. I. (1998). Positional cloning of the zebrafish sauternes gene: a model for congenital sideroblastic anaemia. *Nat Genet* **20**, 244-50.

Brunet, A., Bonni, A., Zigmond, M. J., Lin, M. Z., Juo, P., Hu, L. S., Anderson, M. J., Arden, K. C., Blenis, J. and Greenberg, M. E. (1999). Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857-68.

Brunet, A., Datta, S. R. and Greenberg, M. E. (2001a). Transcription-dependent and -independent control of neuronal survival by the PI3K-Akt signaling pathway. *Curr Opin Neurobiol* **11**, 297-305.

Brunet, A., Park, J., Tran, H., Hu, L. S., Hemmings, B. A. and Greenberg, M. E. (2001b). Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). *Mol Cell Biol* **21**, 952-65.

Clark, M. D., Hennig, S., Herwig, R., Clifton, S. W., Marra, M. A., Lehrach, H., Johnson, S. L. and Group t, W. (2001). An oligonucleotide fingerprint normalized and expressed sequence tag characterized zebrafish cDNA library. *Genome Res* **11**, 1594-602.

Clark, R. F. and Goate, A. M. (1993). Molecular genetics of Alzheimer's disease. *Arch Neurol* **50**, 1164-72.

Collins, F. S. (1990). Identifying human disease genes by positional cloning. *Harvey Lect* **86**, 149-64.

Collins, F. S. (1991). Of needles and haystacks: finding human disease genes by positional cloning. *Clin Res* **39**, 615-23.

Collins, F. S. (1992). Positional cloning: let's not call it reverse anymore. *Nat Genet* **1**, 3-6.

Cook, J. R., Dehner, L. P., Collins, M. H., Ma, Z., Morris, S. W., Coffin, C. M. and Hill, D. A. (2001). Anaplastic lymphoma kinase (ALK) expression in the inflammatory myofibroblastic tumor: a comparative immunohistochemical study. *Am J Surg Pathol* **25**, 1364-71.

Daly, N. L., Djordjevic, J. T., Kroon, P. A. and Smith, R. (1995). Three-dimensional structure of the second cysteine-rich repeat from the human low-density lipoprotein receptor. *Biochemistry* **34**, 14474-81.

Datta, S. R., Brunet, A. and Greenberg, M. E. (1999). Cellular survival: a play in three Akts. *Genes Dev* **13**, 2905-27.

Davidson, A. J., Postlethwait, J. H., Yan, Y. L., Beier, D. R., van Doren, C., Foernzler, D., Celeste, A. J., Crosier, K. E. and Crosier, P. S. (1999). Isolation of zebrafish gdf7 and comparative genetic mapping of genes belonging to the growth/differentiation factor 5, 6, 7 subgroup of the TGF-beta superfamily. *Genome Res* **9**, 121-9.

de Martino, S., Yan, Y. L., Jowett, T., Postlethwait, J. H., Varga, Z. M., Ashworth, A. and Austin, C. A. (2000). Expression of sox11 gene duplicates in zebrafish suggests

the reciprocal loss of ancestral gene expression patterns in development. *Dev Dyn* **217**, 279-92.

Denton , E.J. and Nicol, J.A.C. (1966). A survey of reflectivity in silvery teleosts. *J. mar. Biol. Ass.* U.K. **46**, 685-722.

Denton, E.J. and Rowe, D.M. (1994). Reflectivity communication between fish, with special reference to the greater sand eel, *Hyperoplus lanceolatus. Phil. trans. R. Soc. Lond.* B **182,** 145-158.

Denton, E.J., Gilpin-Brown, J.B. and Wright, P.G. (1972). The angular distribution of the light produced by some mesopelagic fish in relation to their camouflage. *Proc. R. Soc. Lond.* B. **182**, 145-158.

Dickinson, M. E., Selleck, M. A., McMahon, A. P. and Bronner-Fraser, M. (1995). Dorsalization of the neural tube by the non-neural ectoderm. *Development* **121**, 2099-106.

Donovan, A., Brownlie, A., Zhou, Y., Shepard, J., Pratt, S. J., Moynihan, J., Paw, B. H., Drejer, A., Barut, B., Zapata, A. et al. (2000). Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter. *Nature* **403**, 776-81.

Dorsky, R. I., Moon, R. T. and Raible, D. W. (1998). Control of neural crest cell fate by the Wnt signalling pathway. *Nature* **396**, 370-3.

Dorsky, R. I., Raible, D. W. and Moon, R. T. (2000). Direct regulation of nacre, a zebrafish MITF homolog required for pigment cell formation, by the Wnt pathway. *Genes Dev* **14**, 158-62.

Douar, A. M., Mosser, J., Sarde, C. O., Lopez, J., Mandel, J. L. and Aubourg, P. (1994). X-linked adrenoleukodystrophy gene: identification of a candidate gene by positional cloning. *Biomed Pharmacother* **48**, 215-8.

Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Malicki, J., Stemple, D. L., Stainier, D. Y., Zwartkruis, F., Abdelilah, S., Rangini, Z. et al. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37-46.

Dubilet, D. (1995). *Light in the sea.* Koln. Verlag GmbH

Dutton, K. A., Pauliny, A., Lopes, S. S., Elworthy, S., Carney, T. J., Rauch, J., Geisler, R., Haffter, P. and Kelsh, R. N. (2001). Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development* **128**, 4113-25.

Du Shane, G.P. (1935). An experimental study of the origin of pigment cells in amphibia. *J. Exp. Zool.* **72**, 1-31.

Duyster, J., Bai, R. Y. and Morris, S. W. (2001). Translocations involving anaplastic lymphoma kinase (ALK). *Oncogene* **20**, 5623-37.

Eisen, J. S. and Weston, J. A. (1993). Development of the neural crest in the zebrafish. *Dev Biol* **159**, 50-9.

Elworthy, S., Lister, J. A., Carney, T. J., Raible, D. W. and Kelsh, R. N. (2003). Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. *Development* **130**, 2809-18.

Epstein, D.J. Vekemans, M. and Gros, P. (1991) *Splotch* (*Sp2H*), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of *Pax-3*. *Cell* **67**, 767-774.

Erickson, C. A. and Perris, R. (1993). The role of cell-cell and cell-matrix interactions in the morphogenesis of the neural crest. *Dev Biol* **159**, 60-74.

Erickson, C. A. and Reedy, M. V. (1998). Neural crest development: the interplay between morphogenesis and cell differentiation. *Curr Top Dev Biol* **40**, 177-209.

Evens, A. M. and Gartenhaus, R. B. (2003). Molecular etiology of mature T-cell non-hodgkins lymphomas. *Front Biosci* **8**, D156-75.

Fantl, W. J., Johnson, D. E. and Williams, L. T. (1993). Signalling by receptor tyrosine kinases. *Annu Rev Biochem* **62**, 453-81.

Fass, D., Blacklow, S., Kim, P. S. and Berger, J. M. (1997). Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. *Nature* **388**, 691-3.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-45.

Fornzler, D., Her, H., Knapik, E. W., Clark, M., Lehrach, H., Postlethwait, J. H., Zon, L. I. and Beier, D. R. (1998). Gene mapping in zebrafish using single-strand conformation polymorphism analysis. *Genomics* **51**, 216-22.

Garcia-Castro, M. I., Marcelle, C. and Bronner-Fraser, M. (2002). Ectodermal Wnt function as a neural crest inducer. *Science* **297**, 848-51.

Gascoyne, R. D., Aoun, P., Wu, D., Chhanabhai, M., Skinnider, B. F., Greiner, T. C., Morris, S. W., Connors, J. M., Vose, J. M., Viswanatha, D. S. et al. (1999). Prognostic significance of anaplastic lymphoma kinase (ALK) protein expression in adults with anaplastic large cell lymphoma. *Blood* **93**, 3913-21.

Gates, M. A., Kim, L., Egan, E. S., Cardozo, T., Sirotkin, H. I., Dougan, S. T., Lashkari, D., Abagyan, R., Schier, A. F. and Talbot, W. S. (1999). A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. *Genome Res* **9**, 334-47.

Gaudino, G., Follenzi, A., Naldini, L., Collesi, C., Santoro, M., Gallo, K. A., Godowski, P. J. and Comoglio, P. M. (1994). RON is a heterodimeric tyrosine kinase receptor activated by the HGF homologue MSP. *Embo J* **13**, 3524-32.

Geisler, R., Rauch, G. J., Baier, H., van Bebber, F., Brobeta, L., Dekens, M. P., Finger, K., Fricke, C., Gates, M. A., Geiger, H. et al. (1999). A radiation hybrid map of the zebrafish genome. *Nat Genet* **23**, 86-9.

Geissler, E. N., Ryan, M. A. and Housman, D. E. (1988). The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. *Cell* **55**, 185-92.

Goldberg, N. S. and Collins, F. S. (1991). The hunt for the neurofibromatosis gene. *Arch Dermatol* **127**, 1705-7.

Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. and Gruss, P. (1991). Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *Embo J* **10**, 1135-47.

Greenland, C., Touriol, C., Chevillard, G., Morris, S. W., Bai, R., Duyster, J., Delsol, G. and Allouche, M. (2001). Expression of the oncogenic NPM-ALK chimeric protein in human lymphoid T-cells inhibits drug-induced, but not Fas-induced apoptosis. *Oncogene* **20**, 7386-97.

Guo, S., Yamaguchi, Y., Schilbach, S., Wada, T., Lee, J., Goddard, A., French, D., Handa, H. and Rosenthal, A. (2000). A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366-9.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P. et al. (1996a). The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development* **123**, 1-36.

Haffter, P. and Nusslein-Volhard, C. (1996). Large scale genetics in a small vertebrate, the zebrafish. *Int J Dev Biol* **40**, 221-7.

Haffter, P., Odenthal, J., Mullins, M. C., Lin, S., Farrell, M. J., Vogelsang, E., Haas, F., Brand, M., vanEeden, F. J. M., FurutaniSeiki, M. et al. (1996b). Mutations affecting pigmentation and shape of the adult zebrafish. *Development Genes and Evolution* **206**, 260-276.

Hemesath, T. J., Price, E. R., Takemoto, C., Badalian, T. and Fisher, D. E. (1998). MAP kinase links the transcription factor Microphthalmia to c-Kit signalling in melanocytes. *Nature* **391**, 298-301.

Henion, P. D. and Weston, J. A. (1997). Timing and pattern of cell fate restrictions in the neural crest lineage. *Development* **124**, 4351-9.

Herbarth, B., Pingault, V., Bondurand, N., Kuhlbrodt, K., Hermans-Borgmeyer, I., Puliti, A., Lemort, N., Goossens, M. and Wegner, M. (1998). Mutation of the Sry-related Sox10 gene in Dominant megacolon, a mouse model for human Hirschsprung disease. *Proc Natl Acad Sci U S A* **95**, 5161-5.

Hernandez, L., Bea, S., Bellosillo, B., Pinyol, M., Falini, B., Carbone, A., Ott, G., Rosenwald, A., Fernandez, A., Pulford, K. et al. (2002). Diversity of genomic breakpoints in TFG-ALK translocations in anaplastic large cell lymphomas: identification of a new TFG-ALK(XL) chimeric gene with transforming activity. *Am J Pathol* **160**, 1487-94.

Hodgkinson, C. A., Moore, K. J., Nakayama, A., Steingrimsson, E., Copeland, N. G., Jenkins, N. A. and Arnheiter, H. (1993). Mutations at the mouse microphthalmia locus

are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**, 395-404.

Hoffman, E. P., Brown R.H., and Kunkel, L.M. (1987). Dystrophin: the protein product of the Duschenne muscular dystrophy locus. *Cell* **51**, 919-928.

Honda, H., Harada, K., Komuro, I., Terasaki, F., Ueno, H., Tanaka, Y., Kawamura, K., Yazaki, Y. and Hirai, H. (1999). Heart-specific activation of LTK results in cardiac hypertrophy, cardiomyocyte degeneration and gene reprogramming in transgenic mice. *Oncogene* **18**, 3821-30.

Horne-Badovinac, S., Lin, D., Waldron, S., Schwarz, M., Mbamalu, G., Pawson, T., Jan, Y., Stainier, D. Y. and Abdelilah-Seyfried, S. (2001). Positional cloning of heart and soul reveals multiple roles for PKC lambda in zebrafish organogenesis. *Curr Biol* **11**, 1492-502.

Horstadius, S. (1950). *The neural crest: Its properties and derivatives in the light of experimental research*. Oxford University Press. London.

Hosoda, K., Hammer, R. E., Richardson, J. A., Baynash, A. G., Cheung, J. C., Giaid, A. and Yanagisawa, M. (1994). Targeted and natural (piebald-lethal) mutations of endothelin-B receptor gene produce megacolon associated with spotted coat color in mice. *Cell* **79**, 1267-76.

Hukriede, N., Fisher, D., Epstein, J., Joly, L., Tellis, P., Zhou, Y., Barbazuk, B., Cox, K., Fenton-Noriega, L., Hersey, C. et al. (2001). The LN54 radiation hybrid map of zebrafish expressed sequences. *Genome Res* **11**, 2127-32.

Inoue, N., Hess, K. D., Moreadith, R. W., Richardson, L. L., Handel, M. A., Watson, M. L. and Zinn, A. R. (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* **8**, 1201-7.

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A. and de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* **6**, 84-9.

Iovine, M. K. and Johnson, S. L. (2002). A genetic, deletion, physical, and human homology map of the long fin region on zebrafish linkage group 2. *Genomics* **79**, 756-9.

Iwahara, T., Fujimoto, J., Wen, D., Cupples, R., Bucay, N., Arakawa, T., Mori, S., Ratzkin, B. and Yamamoto, T. (1997). Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* **14**, 439-49.

Jessen, J. R., Topczewski, J., Bingham, S., Sepich, D. S., Marlow, F., Chandrasekhar, A. and Solnica-Krezel, L. (2002). Zebrafish trilobite identifies new roles for Strabismus in gastrulation and neuronal movements. *Nat Cell Biol* **4**, 610-5.

Jiang, Y. P., Wang, H., D'Eustachio, P., Musacchio, J. M., Schlessinger, J. and Sap, J. (1993). Cloning and characterization of R-PTP-kappa, a new member of the receptor

protein tyrosine phosphatase family with a proteolytically cleaved cellular adhesion molecule-like extracellular region. *Mol Cell Biol* **13**, 2942-51.

Karlen, S. and Rebagliati, M. (2001). A morpholino phenocopy of the cyclops mutation. *Genesis* **30**, 126-8.

Kelly, P. D., Chu, F., Woods, I. G., Ngo-Hazelett, P., Cardozo, T., Huang, H., Kimm, F., Liao, L., Yan, Y. L., Zhou, Y. et al. (2000). Genetic linkage mapping of zebrafish genes and ESTs. *Genome Res* **10**, 558-67.

Kelsh, R. N. and Raible, D. W. (2002). Specification of zebrafish neural crest. *Results Probl Cell Differ* **40**, 216-36.

Kelsh, R. N. and Eisen, J. S. (2000). The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives. *Development* **127**, 515-25.

Kelsh, R. N., Dutton, K., Medlin, J. and Eisen, J. S. (2000a). Expression of zebrafish fkd6 in neural crest-derived glia. *Mech Dev* **93**, 161-4.

Kelsh, R. N., Schmid, B. and Eisen, J. S. (2000b). Genetic analysis of melanophore development in zebrafish embryos. *Dev Biol* **225**, 277-93.

Kelsh, R. N., Brand, M., Jiang, Y. J., Heisenberg, C. P., Lin, S., Haffter, P., Odenthal, J., Mullins, M. C., vanEeden, F. J. M., FurutaniSeiki, M. et al. (1996). Zebrafish

pigmentation mutations and the processes of neural crest development. *Development* **123**, 369-389.

Kessel, M. and Gruss, P. (1990). Murine developmental control genes. *Science* **249**, 374-9.

Kim, J., Lo, L., Dormand, E. and Anderson, D. J. (2003). SOX10 maintains multipotency and inhibits neuronal differentiation of neural crest stem cells. *Neuron* **38**, 17-31.

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253-310.

Knapik, E. W., Goodman, A., Atkinson, O. S., Roberts, C. T., Shiozawa, M., Sim, C. U., Weksler-Zangen, S., Trolliet, M. R., Futrell, C., Innes, B. A. et al. (1996). A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphisms. *Development* **123**, 451-60.

Knapik, E. W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M. C. and Jacob, H. J. (1998). A microsatellite genetic linkage map for zebrafish (Danio rerio). *Nat Genet* **18**, 338-43.

Kos, R., Reedy, M. V., Johnson, R. L. and Erickson, C. A. (2001). The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. *Development* **128**, 1467-79.

Kozutsumi, H., Toyoshima, H., Hagiwara, K., Furuya, A., Mioh, H., Hanai, N., Yazaki, Y. and Hirai, H. (1993). Identification of the human Itk gene product in placenta and hematopoietic cell lines. *Biochem Biophys Res Commun* **190**, 674-9.

Kupperman, E., An, S., Osborne, N., Waldron, S. and Stainier, D. Y. (2000). A sphingosine-1-phosphate receptor regulates cell migration during vertebrate heart development. *Nature* **406**, 192-5.

Kuroda, H., Mandai, M., Konishi, I., Yura, Y., Tsuruta, Y., Hamid, A. A., Nanbu, K., Matsushita, K. and Mori, T. (1998). Human chorionic gonadotropin (hCG) inhibits cisplatin-induced apoptosis in ovarian cancer cells: possible role of up-regulation of insulin-like growth factor-1 by hCG. *Int J Cancer* **76**, 571-8.

Kwok, C., Critcher, R. and Schmitt, K. (1999). Construction and characterization of zebrafish whole genome radiation hybrids. *Methods Cell Biol* **60**, 287-302.

LaBonne, C. and Bronner-Fraser, M. (1998). Neural crest induction in Xenopus: evidence for a two-signal model. *Development* **125**, 2403-14.

LaBonne, C. and Bronner-Fraser, M. (2000). Snail-related transcriptional repressors are required in Xenopus for both the induction of the neural crest and its subsequent migration. *Dev Biol* **221**, 195-205.

Ladanyi, M. (2000). Aberrant ALK tyrosine kinase signaling. Different cellular lineages, common oncogenic mechanisms. *Am J Pathol* **157**, 341-5.

Land, M. F. (1972). The physics and biology of animal reflectors. *Prog Biophys Mol Biol* **24**, 75-106.

Le Douarin, N.M. and Kalcheim C. (1999). The neural Crest: Cambridge: Cambridge University Press.

Le Douarin, N.M. , Ziller, C. and Couly G.F. (1993). Patterning of neural crest derivatives in the avian embryo: *in vivo* and *in vitro* studies. *Dev. Biol.* **159**, 24-49.

Liem, K. F., Jr., Tremml, G., Roelink, H. and Jessell, T. M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**, 969-79.

Lister, J. A., Close, J. and Raible, D. W. (2001). Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. *Dev Biol* **237**, 333-44.

Lister, J. A., Robertson, C. P., Lepage, T., Johnson, S. L. and Raible, D. W. (1999). nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development* **126**, 3757-67.

Liu, J. P. and Jessell, T. M. (1998). A role for rhoB in the delamination of neural crest cells from the dorsal neural tube. *Development* **125**, 5055-67.

Loren, C. E., Scully, A., Grabbe, C., Edeen, P. T., Thomas, J., McKeown, M., Hunter, T. and Palmer, R. H. (2001). Identification and characterization of DAlk: a novel

Drosophila melanogaster RTK which drives ERK activation in vivo. *Genes Cells* **6**, 531-44.

Lyons, S. E., Lawson, N. D., Lei, L., Bennett, P. E., Weinstein, B. M. and Liu, P. P. (2002). A nonsense mutation in zebrafish gata1 causes the bloodless phenotype in vlad tepes. *Proc Natl Acad Sci U S A* **99**, 5454-9.

Maitre, B., Clement, A., Williams, M. C. and Brody, J. S. (1995). Expression of insulin-like growth factor receptors 1 and 2 in the developing lung and their relation to epithelial cell differentiation. *Am J Respir Cell Mol Biol* **13**, 262-70.

Marusich, M. F., Furneaux, H. M., Henion, P. D. and Weston, J. A. (1994). Hu neuronal proteins are expressed in proliferating neurogenic cells. *J Neurobiol* **25**, 143-55.

Michikawa, M., Kikuchi, S., Muramatsu, H., Muramatsu, T. and Kim, S. U. (1993). Retinoic acid responsive gene product, midkine, has neurotrophic functions for mouse spinal cord and dorsal root ganglion neurons in culture. *J Neurosci Res* **35**, 530-9.

Miller, C. T. and Kimmel, C. B. (2001). Morpholino phenocopies of endothelin 1 (sucker) and other anterior arch class mutations. *Genesis* **30**, 186-7.

Moase, C. E. and Trasler, D. G. (1990). Delayed neural crest cell emigration from Sp and Spd mouse neural tube explants. *Teratology* **42**, 171-82.

Moens, C. B., Cordes, S. P., Giorgianni, M. W., Barsh, G. S. and Kimmel, C. B. (1998). Equivalence in the genetic control of hindbrain segmentation in fish and mouse. *Development* **125**, 381-91.

Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L. and Look, A. T. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263**, 1281-4.

Morris, S. W., Naeve, C., Mathew, P., James, P. L., Kirstein, M. N., Cui, X. and Witte, D. P. (1997). ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene* **14**, 2175-88.

Mullins, M. C., Hammerschmidt, M., Haffter, P. and Nusslein-Volhard, C. (1994). Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol* **4**, 189-202.

Muramatsu, T. (2002). Midkine and pleiotrophin: two related proteins involved in development, survival, inflammation and tumorigenesis. *J Biochem (Tokyo)* **132**, 359-71.

Nakagawa, S. and Takeichi, M. (1995). Neural crest cell-cell adhesion controlled by sequential and subpopulation-specific expression of novel cadherins. *Development* **121**, 1321-32.

Nechiporuk, A., Finney, J. E., Keating, M. T. and Johnson, S. L. (1999). Assessment of polymorphism in zebrafish mapping strains. *Genome Res* **9**, 1231-8.

Nelson, D. L. (1990). Current methods for YAC clone characterization. *Genet Anal Tech Appl* **7**, 100-6.

Newgreen, D. and Gibbins, I. (1982). Factors controlling the time of onset of the migration of neural crest cells in the fowl embryo. *Cell Tissue Res* **224**, 145-60.

Newgreen, D. F. and Gooday, D. (1985). Control of the onset of migration of neural crest cells in avian embryos. Role of Ca++-dependent cell adhesions. *Cell Tissue Res* **239**, 329-36.

Nieto, M. A. (2001). The early steps of neural crest development. *Mech Dev* **105**, 27-35.

Nieto, M. A., Sargent, M. G., Wilkinson, D. G. and Cooke, J. (1994). Control of cell behavior during vertebrate development by Slug, a zinc finger gene. *Science* **264**, 835-9.

Odenthal, J. and Nusslein-Volhard, C. (1998). fork head domain genes in zebrafish. *Dev Genes Evol* **208**, 245-58.

Odenthal, J., Rossnagel, K., Haffter, P., Kelsh, R. N., Vogelsang, E., Brand, M., van Eeden, F. J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M. et al. (1996). Mutations affecting xanthophore pigmentation in the zebrafish, Danio rerio. *Development* **123**, 391-8.

Ota, T. and Amemiya, C. T. (1996). A nonradioactive method for improved restriction analysis and fingerprinting of large P1 artificial chromosome clones. *Genet Anal* **12**, 173-8.

Parichy, D. M., Mellgren, E. M., Rawls, J. F., Lopes, S. S., Kelsh, R. N. and Johnson, S. L. (2000a). Mutational analysis of endothelin receptor b1 (rose) during neural crest and pigment pattern development in the zebrafish Danio rerio. *Dev Biol* **227**, 294-306.

Parichy, D. M., Ransom, D. G., Paw, B., Zon, L. I. and Johnson, S. L. (2000b). An orthologue of the kit-related gene fms is required for development of neural crest-derived xanthophores and a subpopulation of adult melanocytes in the zebrafish, Danio rerio. *Development* **127**, 3031-44.

Parichy, D. M., Rawls, J. F., Pratt, S. J., Whitfield, T. T. and Johnson, S. L. (1999). Zebrafish sparse corresponds to an orthologue of c-kit and is required for the morphogenesis of a subpopulation of melanocytes, but is not essential for hematopoiesis or primordial germ cell development. *Development* **126**, 3425-3436.

Parsons, M. J., Pollard, S. M., Saude, L., Feldman, B., Coutinho, P., Hirst, E. M. and Stemple, D. L. (2002). Zebrafish mutants identify an essential role for laminins in notochord formation. *Development* **129**, 3137-46.

Pavan, W. J. and Tilghman, S. M. (1994). Piebald lethal (sl) acts early to disrupt the development of neural crest-derived melanocytes. *Proc Natl Acad Sci U S A* **91**, 7159-63.

Payne, T. L., Postlethwait, J. H. and Yelick, P. C. (2001). Functional characterization and genetic mapping of alk8. *Mech Dev* **100**, 275-89.

Pera, E. M., Wessely, O., Li, S. Y. and De Robertis, E. M. (2001). Neural and head induction by insulin-like growth factor signals. *Dev Cell* **1**, 655-65.

Pingault, V., Bondurand, N., Kuhlbrodt, K., Goerich, D. E., Prehu, M. O., Puliti, A., Herbarth, B., Hermans-Borgmeyer, I., Legius, E., Matthijs, G. et al. (1998). SOX10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet* **18**, 171-3.

Poss, K. D., Nechiporuk, A., Hillam, A. M., Johnson, S. L. and Keating, M. T. (2002). Mps1 defines a proximal blastemal proliferative compartment essential for zebrafish fin regeneration. *Development* **129**, 5141-9.

Postlethwait, J., Amores, A., Force, A. and Yan, Y. L. (1999). The zebrafish genome. *Methods Cell Biol* **60**, 149-63.

Postlethwait, J. H., Yan, Y. L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., Gong, Z. et al. (1998). Vertebrate genome evolution and the zebrafish gene map. *Nat Genet* **18**, 345-9.

Powers, C., Aigner, A., Stoica, G. E., McDonnell, K. and Wellstein, A. (2002). Pleiotrophin signaling through anaplastic lymphoma kinase is rate-limiting for glioblastoma growth. *J Biol Chem* **277**, 14153-8.

Pulford, K., Morris, S. W. and Mason, D. Y. (2001). Anaplastic lymphoma kinase proteins and malignancy. *Curr Opin Hematol* **8**, 231-6.

Raible, D. W. and Eisen, J. S. (1994). Restriction of neural crest cell fate in the trunk of the embryonic zebrafish. *Development* **120**, 495-503.

Raible, D. W. and Eisen, J. S. (1996). Regulative interactions in zebrafish neural crest. *Development* **122**, 501-7.

Ransom, D. G. and Zon, L. I. (1999). Mapping zebrafish mutations by AFLP. *Methods Cell Biol* **60**, 195-211.

Ranta, S., Zhang, Y., Ross, B., Takkunen, E., Hirvasniemi, A., de la Chapelle, A., Gilliam, T. C. and Lehesjoki, A. E. (2000). Positional cloning and characterisation of the human DLGAP2 gene and its exclusion in progressive epilepsy with mental retardation. *Eur J Hum Genet* **8**, 381-4.

Rauch, G. J. G., M. and Haffter, P. (1997). A polymorphic zebrafish line for genetic mapping using SSLPs on high-percentage agarose gels. *Elsevier Trends Journals Technical Tips Online*.

Rawles, M.E. (1947). Some observations on the developmental properties of the presumptive himd-limb area of the chick. *Anat. Rec.* **99**, 648-649.

Rawles, M.E., (1948). Origin of melanophores and their role in development of color patterns in vertebrates. *Physiol. Rev.* **28**, 383-408.

Robinson, D. R., Wu, Y. M. and Lin, S. F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene* **19**, 5548-57.

Rottbauer, W., Baker, K., Wo, Z. G., Mohideen, M. A., Cantiello, H. F. and Fishman, M. C. (2001). Growth and function of the embryonic heart depend upon the cardiac-specific L-type calcium channel alpha1 subunit. *Dev Cell* **1**, 265-75.

Russel, W.L. (1947). *Splotch*, a new mutation in the house mouse *Mus musculus. Genetics.* **32,** 107-111.

Sambrook, J., Fritsch,E.F., Maniatis, T. (1989). Molecular cloning. A laboratory manual. New York: Cold Spring Harbour Laboratory Press.

Santiago, A. and Erickson, C. A. (2002). Ephrin-B ligands play a dual role in the control of neural crest cell migration. *Development* **129**, 3621-32.

Schilling, T. F. and Kimmel, C. B. (1994). Segment and cell type lineage restrictions during pharyngeal arch development in the zebrafish embryo. *Development* **120**, 483-94.

Sela-Donenfeld, D. and Kalcheim, C. (1999). Regulation of the onset of neural crest migration by coordinated activity of BMP4 and Noggin in the dorsal neural tube. *Development* **126**, 4749-62.

Selleck, M. A. and Bronner-Fraser, M. (1995). Origins of the avian neural crest: the role of neural plate-epidermal interactions. *Development* **121**, 525-38.

Serbedzija, G. N. and McMahon, A. P. (1997). Analysis of neural crest cell migration in Splotch mice using a neural crest-specific LacZ reporter. *Dev Biol* **185**, 139-47.

Shafizadeh, E., Paw, B. H., Foott, H., Liao, E. C., Barut, B. A., Cope, J. J., Zon, L. I. and Lin, S. (2002). Characterization of zebrafish merlot/chablis as non-mammalian vertebrate models for severe congenital anemia due to protein 4.1 deficiency. *Development* **129**, 4359-70.

Shimoda, N., Knapik, E. W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H. and Fishman, M. C. (1999). Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**, 219-32.

Shin, M. K., LevorseJ.M., Ingram, R.S., Tilghman, S.M. (1999). The temporal requirement for endothelin receptor-B signalling during neural crest development. *Nature* **402**, 496-501.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A* **89**, 8794-7.

Sieber-Blum, M. (1990). Mechanisms of neural crest diversification. *Comments Dev.Neurobiol.* **1**, 225-249.

Silver. (1995). Mouse genetics. Cambridge University Press.

Snijders, A. J., Haase, V. H. and Bernards, A. (1993). Four tissue-specific mouse ltk mRNAs predict tyrosine kinases that differ upstream of their transmembrane segment. *Oncogene* **8**, 27-35.

Southard-Smith, E. M., Kos, L. and Pavan, W. J. (1998). Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat Genet* **18**, 60-4.

Stemple, D. L. and Anderson, D. J. (1992). Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell* **71**, 973-85.

Stickney, H. L., Schmutz, J., Woods, I. G., Holtzer, C. C., Dickson, M. C., Kelly, P. D., Myers, R. M. and Talbot, W. S. (2002). Rapid Mapping of Zebrafish Mutations With SNPs and Oligonucleotide Microarrays. *Genome Res* **12**, 1929-34.

Stoica, G. E., Kuo, A., Aigner, A., Sunitha, I., Souttou, B., Malerczyk, C., Caughey, D. J., Wen, D., Karavanov, A., Riegel, A. T. et al. (2001). Identification of anaplastic lymphoma kinase as a receptor for the growth factor pleiotrophin. *J Biol Chem* **276**, 16772-9.

Stoica, G. E., Kuo, A., Powers, C., Bowden, E. T., Sale, E. B., Riegel, A. T. and Wellstein, A. (2002). Midkine binds to anaplastic lymphoma kinase (ALK) and acts as a growth factor for different cell types. *J Biol Chem* **277**, 35990-8.

Talbot, W. S. and Schier, A. F. (1999). Positional cloning of mutated zebrafish genes. *Methods Cell Biol* **60**, 259-86.

Topczewski, J., Sepich, D. S., Myers, D. C., Walker, C., Amores, A., Lele, Z., Hammerschmidt, M., Postlethwait, J. and Solnica-Krezel, L. (2001). The zebrafish glypican knypek controls cell polarity during gastrulation movements of convergent extension. *Dev Cell* **1**, 251-64.

Weston, J. A. (1982). Neural crest cell development. *Prog Clin Biol Res* **85**, 359-79.

Woods, I. G., Kelly, P. D., Chu, F., Ngo-Hazelett, P., Yan, Y. L., Huang, H., Postlethwait, J. H., Talbot, W. S. (2000). A comparative map of the zebrafish genome. *Genome Res* **10**, 1903-14

Winkler, C. and Moon, R. T. (2001). Zebrafish mdk2, a novel secreted midkine, participates in posterior neurogenesis. *Dev Biol* **229**, 102-18.

Winkler, C., Schafer, M., Duschl, J., Schartl, M. and Volff, J. N. (2003). Functional divergence of two zebrafish midkine growth factors following fish-specific gene duplication. *Genome Res* **13**, 1067-81.

Xu, X., Meiler, S. E., Zhong, T. P., Mohideen, M., Crossley, D. A., Burggren, W. W. and Fishman, M. C. (2002). Cardiomyopathy in zebrafish due to mutation in an alternatively spliced exon of titin. *Nat Genet* **30**, 205-9.

Yan, Y. L., Talbot, W. S., Egan, E. S. and Postlethwait, J. H. (1998). Mutant rescue by BAC clone injection in zebrafish. *Genomics* **50**, 287-289.

Yanagisawa, M. (1994). The endothelin system. A new target for therapeutical intervention. *Circulation* **89**, 1320-1322.

Yarden, Y. and Ullrich, A. (1988). Growth factor receptor tyrosine kinases. *Annu Rev Biochem* **57**, 443-78.

Yoshida, H., Hayashi, S., Shultz, L. D., Yamamura, K., Nishikawa, S. and Kunisada, T. (1996). Neural and skin cell-specific expression pattern conferred by steel factor regulatory sequence in transgenic mice. *Dev Dyn* **207**, 222-32.

Yoshida, Y., Goto, M., Tsutsui, J., Ozawa, M., Sato, E., Osame, M. and Muramatsu, T. (1995). Midkine is present in the early stage of cerebral infarct. *Brain Res Dev Brain Res* **85**, 25-30.

Zamo, A., Chiarle, R., Piva, R., Howes, J., Fan, Y., Chilosi, M., Levy, D. E. and Inghirami, G. (2002). Anaplastic lymphoma kinase (ALK) activates Stat3 and protects hematopoietic cells from cell death. *Oncogene* **21**, 1038-47.

Zhang, J., Talbot, W. S. and Schier, A. F. (1998). Positional cloning identifies zebrafish one-eyed pinhead as a permissive EGF-related ligand required during gastrulation. *Cell* **92**, 241-51.

Zwick, E., Bange, J. and Ullrich, A. (2001). Receptor tyrosine kinase signalling as a target for cancer intervention strategies. *Endocr Relat Cancer* **8**, 161-73.